

---

# Semiempirical Treatment of Electrostatic Potentials and Partial Charges in Combined Quantum Mechanical and Molecular Mechanical Approaches

---

**DIRK BAKOWIES and WALTER THIEL\***

*Organisch-Chemisches Institut, Universität Zürich, Winterthurerstr. 190, CH-8057 Zürich, Switzerland*

*Received 2 May 1995; accepted 16 July 1995*

---

## ABSTRACT

A semiempirical treatment of electrostatic potentials and partial charges is presented. These are the basic components needed for the evaluation of electrostatic interaction energies in combined quantum mechanical and molecular mechanical approaches. The procedure to compute electrostatic potentials uses AM1 and MNDO wave functions and is based on one previously suggested by Ford and Wang. It retains the NDDO approximation and is thus both easy to implement and computationally efficient. Partial atomic charges are derived from a semiempirical charge equilibration model, which is based on the principle of electronegativity equalization. Large sets of *ab initio* restricted Hartree-Fock (RHF/6-31G\*) reference data have been used to calibrate the semiempirical models. Applying the final parameters (C, H, N, O), the *ab initio* electrostatic potentials are reproduced with an average accuracy of 20% (AM1) and 25% (MNDO), respectively, and the *ab initio* potential derived charges normally to within 0.1 e. In most cases our parameterized models are more accurate than the much more expensive quasi *ab initio* techniques, which employ deorthogonalized semiempirical wave functions and have generally been preferred in previous applications. © 1996 by John Wiley & Sons, Inc.

\*Author to whom all correspondence should be addressed.

## Introduction

There is increasing interest in applying combined quantum mechanical (QM) and molecular mechanical (MM) approaches to study chemical systems which are too large for rigorous quantum chemical treatments. A number of such QM/MM schemes have been proposed<sup>1-7</sup> which are designed for investigations in biochemistry<sup>8,9</sup> and for simulations of solvent effects.<sup>10-19</sup> We have recently developed semiempirical approaches of varying complexity<sup>20,21</sup> that combine the quantum chemical methods MNDO<sup>22</sup> and AM1<sup>23</sup> with the MM3<sup>24-26</sup> force field. Particular attention has been paid to the consistent semiempirical treatment of electrostatic and induction interactions involving both the QM and MM regions of the system under study and a proper calibration of the model parameters against reliable *ab initio* (RHF/6-31G\*) reference data.

This article focuses on the development and parameterization of the semiempirical models involved. Key quantities of the approach are the electrostatic potentials  $\langle \Phi^I \rangle$  generated by the QM part of the supermolecule at the positions of the MM atoms  $I$  and the atomic point (or partial) charges  $q_I$ , which set up the electrostatic potential of the MM region. The Coulomb energy between the QM and MM parts of the system may then simply be expressed as a sum of electrostatic interactions:

$$E^{\text{coul}} = \sum_I q_I \langle \Phi^I \rangle \quad (1)$$

In LCAO approximation (linear combination of atomic orbitals) the electrostatic potential is given by

$$\langle \Phi^I \rangle = - \sum_{\mu\nu} P_{\mu\nu} V_{\mu\nu}^I + \sum_A Z_A V^{AI} \quad (2)$$

where  $P_{\mu\nu}$  and  $Z_A$  are density matrix elements and nuclear charges, respectively.  $V_{\mu\nu}^I$  and  $V^{AI}$  denote nuclear attraction integrals and Coulomb repulsion terms which describe the interaction between a unit test charge and an electron or a nucleus, respectively. The negative gradient of the electrostatic potential defines the electric field which is needed for the evaluation of induction energies:

$$\langle F_{\alpha}^I \rangle = \sum_{\mu\nu} P_{\mu\nu} \nabla_{\alpha} V_{\mu\nu}^I - \sum_A Z_A \nabla_{\alpha} V^{AI} \quad (3)$$

Following eqs. (2) and (3), the computation of potentials and fields is straightforward within the framework of *ab initio* methodology. In semiempirical theory the situation is less clear. The neglect of some two-center and all three- and four-center integrals in an (assumed) orthogonal AO basis calls for the need to represent the remaining integrals and the core-core repulsions by suitable parametric formulas calibrated against experimental results. Hence there is no obvious reason for an exact correspondence between one-electron integrals and Coulomb terms describing the interaction between atoms and those entering the formulas for electrostatic potentials and electric fields [eqs. (2) and (3)]. In semiempirical theory the two-center one-electron integrals and the core-core repulsion terms depend on empirical parameters for both interacting atoms, whereas neither  $V_{\mu\nu}^I$  nor  $V^{AI}$  should contain parameters connected with the index  $I$  (denoting the location where the potential is measured), since this would introduce inconsistencies with the definition of the electrostatic potential. This may explain the shortcomings of the early purely semiempirical approaches to calculate electrostatic potentials employing the same formulas and parameters as for the interaction between atoms.<sup>27</sup>

Based on the pioneering work of Giessner-Prettre and Pullman,<sup>28</sup> the *ab initio* evaluation of the one-electron integrals  $V_{\mu\nu}^I$  in a deorthogonalized basis has found wide appreciation.<sup>27,29</sup> The considerable computational requirements of this conceptually simple quasi *ab initio* approach, however, have stimulated further research on alternative semiempirical procedures,<sup>30-32</sup> with only moderate success.<sup>29</sup> In a promising recent development, Rauhut and Clark have presented a model for AM1 and PM3 wave functions which approximates the electrostatic potential by point charges at the centers of charge density given by the natural atomic orbitals (for an *sp* basis set).<sup>33,34</sup>

We use an alternative purely semiempirical scheme for MNDO and AM1 wave functions similar to one introduced by Ford and Wang,<sup>35</sup> who replaced the usual expressions for the one-electron integrals and nuclear repulsion functions in eqs. (2) and (3) by semiempirical formulas which were parameterized to reproduce *ab initio* RHF/6-31G\* results. Their parameterization was tailored to provide an optimal representation of electrostatic potentials in the vicinities of the minima associates with organic functional groups. We have reparameterized this model to provide a more evenly balanced account of the electrostatic potential re-

quired in the present application. A detailed comparison will be given in this article, with special emphasis on the applicability in combined QM/MM approaches.

The partial charges  $q_i$  entering eq. (1) require further consideration. Neither in semiempirical nor in *ab initio* QM/MM approaches can they be defined unambiguously, as they do not refer to the expectation value of any quantum mechanical operator. It is clear, however, that for our purposes they should be related to the electrostatic potential of the molecule under consideration. This would suggest the use of potential derived (PD)<sup>36</sup> charges, which are, however, impractical in a combined QM/MM method since their evaluation requires a quantum chemical calculation for the classical part of the supermolecule. An empirical scheme able to approximate PD charges is therefore more appropriate. The principle of electronegativity equalization, suggested by Sanderson<sup>37-39</sup> and confirmed within density functional theory by Parr et al.,<sup>40</sup> serves as the physical basis for simple charge equilibration models proposed by several groups.<sup>41-43</sup> We present a semiempirical reformulation of the approach suggested by Rappé and Goddard,<sup>43</sup> in which the model parameters are adjusted to reproduce *ab initio* RHF/6-31G\*/PD charges as closely as possible.

This article is organized as follows: The first section describes some general computational aspects. The second section deals with the computation of electrostatic potentials (and electric fields), beginning with the basis ideas, continuing with specific details of the parameterization, and ending with a critical discussion of the results obtained. The third section presents the work concerning the partial charges in a similar fashion. In the Appendix, the analytical gradients of the geometry-dependent model charges are given, since they are needed for an efficient implementation of QM/MM approaches which employ the proposed charge model.

### Computational Details

All parameterizations have been carried out using a computer program specifically designed for this purpose. It is based on PP91,<sup>44</sup> a code for semiempirical parameterizations developed in our laboratory, and uses nonlinear optimization techniques as implemented by Weiner<sup>45</sup> and Kolb.<sup>46</sup> *Ab initio* reference data have been obtained for 24 representative molecules, covering a large range of

typical bonding situations in organic molecules built from H, C, N, and O (see Table I). RHF/6-31G\* geometries have been used throughout, with the exception of molecules 15 and 20 (hydrogen peroxide and *trans*-2,3-diazabut-2-en), where MP2/6-31G\* geometry optimizations have been performed to correct for known deficiencies of the one-determinantal approximation.<sup>47</sup>

Reference data for the parameterization of the electrostatic potentials (and electric fields) and the generation of PD charges have been collected for selected points on the van der Waals surfaces of the reference molecules. A chosen surface is defined by a distance of  $f_{\text{vdW}} \cdot r_{\text{rdW}}^A$  between a point

**TABLE I.**  
Reference Molecules.

Molecule	Number of Reference Points on van der Waals Surfaces <sup>a-c</sup>		
	$f_{\text{vdW}} =$ 0.6 ... 1.2	$f_{\text{vdW}} =$ 1.4 ... 2.0	$f_{\text{vdW}} =$ 3.0
1 Aziridine	217	492	300
2 C <sub>2</sub> H <sub>2</sub>	130	420	272
3 C <sub>2</sub> H <sub>4</sub>	172	452	274
4 C <sub>2</sub> H <sub>6</sub>	202	474	272
5 C <sub>6</sub> H <sub>6</sub>	318	702	350
6 CH <sub>2</sub> NH	155	402	276
7 CH <sub>2</sub> O	129	370	247
8 CH <sub>3</sub> CHO	199	494	289
9 CH <sub>3</sub> CN	183	490	318
10 CH <sub>3</sub> NH <sub>2</sub>	192	451	291
11 CH <sub>3</sub> NO <sub>2</sub>	235	507	319
12 CH <sub>4</sub>	142	361	244
13 CO	103	334	245
14 CO <sub>2</sub>	128	375	244
15 H <sub>2</sub> O <sub>2</sub>	134	372	251
16 HCOOCH <sub>3</sub>	242	537	333
17 HCOOH	163	441	281
18 CH <sub>3</sub> OH	175	424	270
19 N <sub>2</sub> H <sub>4</sub>	163	427	287
20 CH <sub>3</sub> NNCH <sub>3</sub> ( <i>trans</i> )	253	587	340
21 NH <sub>2</sub> CHO	179	465	294
22 NH <sub>3</sub>	119	326	246
23 Oxirane	191	459	296
24 H <sub>2</sub> O	92	289	212
TOTAL	4216	10651	6751

<sup>a</sup>The numbers refer to the statistical evaluations reported in Tables III, IV, and VII. In the parameterizations, a reduced number of reference points has been used. See text.

<sup>b</sup>The first two columns comprise four molecular surfaces each ( $f_{\text{vdW}} = 0.6, 0.8, 1.0, 1.2$ , and  $f_{\text{vdW}} = 1.4, 1.6, 1.8, 2.0$ , respectively).

<sup>c</sup>Potential derived charges have been obtained using the points counted in the second column.

on the surface and the atom  $A$ , where  $r_{\text{vdW}}^A$  denotes the van der Waals radius of this atom and  $f_{\text{vdW}}$  denotes a constant factor characteristic of the surface. The points on these molecular surfaces have been obtained using the Connolly algorithm<sup>48</sup> as implemented by Merz<sup>49</sup> in MOPAC 6.0.<sup>50</sup> Standard van der Waals radii (in Å: H 1.2, C 1.5, N 1.5, O 1.4) and point densities ( $1.0 \text{ Å}^{-2}$ ) have been applied.<sup>49,51</sup> All *ab initio* calculations have been carried out using the Gaussian 92 program,<sup>52</sup> and semiempirical results based on deorthogonalized wave functions have been obtained with MOPAC 6.0.<sup>50</sup>

## The Electrostatic Potential

### GENERAL CONSIDERATIONS

According to intermolecular perturbation theory,<sup>53</sup> the interaction between two distant charge distributions  $X$  and  $Y$  is a simple Coulomb interaction in the first order which may be modeled as the energy of point charges in  $Y$  acting in the electrostatic potential of  $X$  or, equivalently, as the energy of point charges in  $X$  acting in the electrostatic potential of  $Y$ . The induction energy occurring in the second order may be expressed as the sum of the energies of induced dipole moments in  $X$  in the electric field of  $Y$  plus the energy of induced dipole moments in  $Y$  in the electric field of  $X$ . Therefore, both the electrostatic potential and the electric field are suitable quantities to parameterize the one-electron integrals  $V_{\mu\nu}^I$  and the Coulomb terms  $V^{AI}$  occurring in the QM/MM schemes [cf. eqs. (1)–(3) and refs. 20 and 21].

Contrary to most of the previous work on semiempirical electrostatic potentials,<sup>28,29,54–56</sup> we assume an orthogonal AO basis as discussed in the Introduction and express the remaining terms  $V_{\mu_A\mu_A}^I$  and  $V^{AI}$  by semiempirical formulas which are analogous to those in the MNDO formalism<sup>22</sup>:

$$V_{\mu_A\mu_A}^I = (\mu_A \mu_A, s_I s_I) \quad (4)$$

$$V^{AI} = (s_A s_A, s_I s_I) \cdot f'(A, I) \quad (5)$$

The unit test charge  $q_I$  measuring the electrostatic potential is only formally represented by a  $(s_I s_I)$  charge distribution, as will be seen later. The two-center integrals  $(\mu_A \mu_A, s_I s_I)$  and  $(s_A s_A, s_I s_I)$  include Klopman-Ohno<sup>57,58</sup> factors accounting for the damping of classical Coulomb interactions due to overlapping electron densities. This is completely in line with the MNDO approximation,

which introduces two atomic parameters  $\rho_{I_A}^A$  and  $\rho_{I_B}^B$  simulating the damping effect. Assuming a distance  $d_{\lambda_A, \lambda_B}$  between the classical point charges representing the multipole moments of the charge distributions  $(\mu_A \nu_A)$  and  $(\lambda_B \sigma_B)$ , these Klopman-Ohno factors read as follows:

$$f_{KO}^{A, B, I_A, I_B} = \left( d_{\lambda_A, \lambda_B}^2 + (\rho_{I_A}^A + \rho_{I_B}^B)^2 \right)^{-\frac{1}{2}} \quad (6)$$

In the calculation of the electrostatic potential, we treat the unit test charge  $q_I$  as a classical charge ( $\rho_0^I = 0$ ) and therefore only employ Klopman-Ohno factors of the following kind:

$$f_{KO}^{A, I, I_A, 0} = \left( d_{\lambda_A, 1}^2 + \rho_{I_A}^{A^2} \right)^{-\frac{1}{2}} \quad (7)$$

Analogously, we allow the function  $f'(A, I)$  in eq. (5) to depend only on parameters of atom  $A$ . The term  $f'(A, I)$  should decrease monotonically with the distance  $r_{AI}$  between  $A$  and  $I$ , approaching a limit of  $f'(A, I) = 1$  for  $r_{AI} \rightarrow \infty$ .

### PARAMETERIZATIONS

The function  $f'(A, I)$  has to be calibrated with respect to the parameters of atom  $A$ . This optimization problem may be formulated conveniently as the minimization of the sum of squares (SSQ) function, which is defined as

$$\text{SSQ} = \sum_{n=1}^{n_{\text{ref}}^{\Phi}} (\omega^{\Phi} \omega^n)^2 (\Phi^{I_n} - \Phi_{\text{ref}}^{I_n})^2 + \sum_{n=1}^{n_{\text{ref}}^F} (\omega^F \omega^n)^2 (F_{\alpha}^{I_n} - F_{\text{ref}}^{I_n})^2 \quad (8)$$

where  $\Phi_{\text{ref}}^{I_n}$  and  $F_{\text{ref}}^{I_n}$  denote the  $n$ th of a total of  $n_{\text{ref}}^{\Phi}$  and  $n_{\text{ref}}^F$  reference functions (electrostatic potential or electric field component at location  $I_n$ ), respectively. The term  $\omega^n$  is a specific weight factor characteristic of a single reference function, while  $\omega^{\Phi}$  and  $\omega^F$  denote global weights discriminating between electrostatic potentials and electric field components.

The choice of reference data is critical and deserves some comment. Ford and Wang calibrated the function  $f'(A, I)$  using RHF/6-31G\* electrostatic potentials on a  $0.2 \text{ Å}$  grid in, and  $2 \text{ Å}$  above, the plane of the cytosine molecule.<sup>35</sup> To weight more heavily those regions associated with electrostatic potential minima, they elected to exclude all regions of space within about 80% of the van der Waals radii of the atoms. The resulting parameters

were further tailored to the calculation of electrostatic potential minima in a subsequent refinement based on RHF/6-31G\* minima for a small set of reference compounds.<sup>35</sup> For a semiempirical QM/MM scheme, however, such a choice seems to be too limited, since any successful QM/MM application will generally require realistic QM electrostatic potentials over the whole space accessible to the MM part of the molecule. This will predominantly include regions of medium or low electron density far away from the minima of the electrostatic potential.

Accordingly, for each of the 24 reference molecules (see Table I), we have distributed the reference points over nine different van der Waals surfaces. In total, there are  $n_{\text{ref}}^{\Phi} = 21,618$  reference functions for the electrostatic potentials and  $n_{\text{ref}}^F = 64,854$  reference functions for the three field components, all of which have been used in the statistical evaluations. The parameterization process, however, is limited to a much smaller number of reference functions (ca. 3000 for an available computer memory of 100 MB) due to the use of the Levenberg-Marquardt algorithm in our program.<sup>44,46</sup> Hence, the parameterization runs require a considerable reduction of the reference data, which has been specified in detail elsewhere<sup>20</sup> and will therefore be outlined only briefly in the following.

For the optimization of parameters, molecules 8, 13, 14, 16, and 21 have been deleted from the list of reference molecules (see Table I), and only surfaces from the first ( $f_{\text{vdW}} = 0.6 \cdots 1.2$ ) or second ( $f_{\text{vdW}} = 1.4 \cdots 2.0$ ) block have been considered. Symmetry equivalent surface points have been included only once, and the resulting density of points has been reduced according to statistical criteria, which guarantee to yield a representative subset of reference points. Finally, only the electrostatic potentials have been admitted as reference functions ( $\omega^{\Phi} = 1$ ,  $\omega^F = 0$ ), except for the surface with  $f_{\text{vdW}} = 0.8$ , where the electric field components served as additional reference data in some of the parameterizations.

In this way the number of reference functions has been reduced significantly—for example, from 86,472 to 2767 for parameterizations including both the electrostatic potential (for  $f_{\text{vdW}} = 0.6 \cdots 1.2$ ) and the electric field (for  $f_{\text{vdW}} = 0.8$ ). The parameterizations proceeded in the following steps:

1. Parameterization for AM1 wave functions:
  - Definition of a specific protocol by choosing the parametric function  $f'(A, I)$ , the

weight  $\omega^F$ , and the molecular surfaces  $f_{\text{vdW}}$

- Optimization of parameters
  - Statistical evaluation of the results for all 86,472 reference data
  - Preselection of the most appropriate protocols
2. Parameterization for MNDO wave functions:
    - Optimization of parameters for the preselected protocols
  3. Final selection of identical protocols for AM1 and MNDO

The quantities used for the statistical evaluations are defined as follows:

$$\text{rms} = \left( \frac{1}{n_{\text{ref}}^x} \sum_{i=1}^{n_{\text{ref}}^x} (x^i - x_{\text{ref}}^i)^2 \right)^{\frac{1}{2}} \quad (\text{root mean square deviation}) \quad (9a)$$

$$\text{rrms} = \text{rms} \cdot \left( \frac{1}{n_{\text{ref}}^x} \sum_{i=1}^{n_{\text{ref}}^x} x_{\text{ref}}^{i^2} \right)^{-\frac{1}{2}} \quad (\text{relative root mean square deviation}) \quad (9b)$$

$$\Delta_{\pm} = \frac{1}{n_{\text{ref}}^x} \sum_{i=1}^{n_{\text{ref}}^x} (x^i - x_{\text{ref}}^i) \quad (\text{mean deviation}) \quad (9c)$$

$$\Delta_{\text{abs}} = \frac{1}{n_{\text{ref}}^x} \sum_{i=1}^{n_{\text{ref}}^x} |x^i - x_{\text{ref}}^i| \quad (\text{mean absolute deviation}) \quad (9d)$$

$$\Delta_{\text{abs}}^{\pm} = \frac{1}{n_{\text{ref}}^x} \sum_{i=1}^{n_{\text{ref}}^x} |x^i| - |x_{\text{ref}}^i| \quad (\text{mean deviation of absolute values}) \quad (9e)$$

## RESULTS AND DISCUSSION

The determination of PD charges is usually based on electrostatic potentials on surfaces with intermediate distances to the molecule ( $f_{\text{vdW}} = 1.4 \cdots 2.0$ ).<sup>51</sup> These surfaces are not, however, suitable for the generation of reference data in the parameterization of semiempirical potentials, since the investigated functions  $f'(A, I)$  fall off much faster with increasing distance than the Coulomb potential involved in the derivation of PD charges. Preliminary optimizations using points on these surfaces showed a very low sensitivity with regard to the values of the parameters and confirmed this assumption empirically.

Thus, all further parameterizations were limited to reference points near the molecules ( $f_{\text{vdW}} = 0.6 \cdots 1.2$ ). Several functions  $f'(A, I)$  with one, two, or three parameters were studied. Most of the functions were exponentials, and some of them were expansions involving inverse powers of  $r_{AI}$ . Even though the statistics did not show any convincing advantage for one of these types of functions, we refrained from investigating further the expansions in favor of a treatment which matches the MNDO formalism as closely as possible. Normally, the statistical results for one-parametric functions were improved by the addition of one further parameter, while the addition of a third parameter always only showed negligible effects. This observation restricted the choice of functions to exponentials with either one ( $\omega_A$ ) or two ( $\omega_A, \delta_A$ ) parameters:

$$f'_1(A, I) = 1 + \exp(-\omega_A r_{AI}) \quad (10a)$$

$$f'_2(A, I) = 1 + \exp(-\omega_A(r_{AI} - \delta_A)) \quad (10b)$$

For both functions, several parameterization protocols have been tested which only differed by the actual choice of reference data. Here we only briefly comment on our experiences that led to the final decision for a specific parameterization. A more complete discussion may be found elsewhere.<sup>20</sup>

Most noticeably, the effects of specific changes in the parameterization protocol were similar for AM1 and MNDO wave functions. Some general conclusions are thus valid for both semiempirical methods. In those parameterization runs which only include electrostatic potentials as reference functions, the relative errors of the electrostatic potentials and the electric fields are similar for both types of exponentials, at least on surfaces with intermediate or large distances to the molecules. On surfaces near the molecular periphery, the two-parametric function shows slight advantages for the electrostatic potential but is less favorable for electric fields. Minima of electrostatic potentials are normally computed slightly more accurately with the two-parametric alternative, although some exceptions have been noted. The inclusion of electric field components as reference data (for  $f_{\text{vdW}} = 0.8$ ) in the parameterizations leads to improved electric fields only for the one-parametric function. The opposite is true for the two-parametric function, which obviously is too flexible and only shows improved results for those molecular surfaces explicitly considered in the parameterization ( $f_{\text{vdW}} = 0.8$ ). In an overall view, both for AM1 and MNDO, the best one-parametric

[ $f'_1(A, I)$ ,  $\omega^\Phi = 1$ ,  $\omega^F = 1$ ] and the best two-parametric [ $f'_2(A, I)$ ,  $\omega^\Phi + 1$ ,  $\omega^F = 0$ ] functions show results of comparable quality. Therefore, both variants are discussed later in more detail. In the proposed QM/MM models, however, we prefer the one-parametric alternative for reasons of simplicity.<sup>20,21</sup>

Table II lists our preferred parameter sets for AM1 and MNDO. The corresponding results are collected in Tables III to VI. On the whole, the relative errors of the electrostatic potentials and the electric fields are a bit larger for MNDO (Table IV) than for AM1 (Table III). In both cases, the distances  $r_{AI}$  of potential minima are calculated too short in the vicinity of electronegative atoms and too long in the vicinity of  $\pi$  systems, but the errors are larger for MNDO (Tables V and VI). Likewise, the underestimation of potential depths of  $\pi$  systems is more pronounced for MNDO. Furthermore, there is a qualitative error concerning the location of the potential minima of  $\text{H}_2\text{O}$  and  $\text{CH}_2\text{O}$ . While *ab initio* methods predict the minimum to lie in the direction of the lone pairs, all of the tested MNDO variants yield the minimum in the  $\text{C}_2$  axis. The AM1 parameterizations lead to qualitatively correct results but do not satisfy quantitatively since the potential differences between transition state (field gradient with one positive eigenvalue) and minimum are underestimated severely [e.g., AM1,  $f'_1(A, I)$ : 0.2 ( $\text{H}_2\text{O}$ ) and 1.2 ( $\text{CH}_2\text{O}$ ) kcal/mol e; RHF/6-31G\*: 1.6 and 5.6 kcal/mol e].

Table VII shows further statistical results for the final parameter sets. The negative mean deviation and the negative mean deviation of absolute values indicate that the electrostatic potential is usu-

**TABLE II.**  
Final Parameters for the Calculation of  
Electrostatic Potentials.<sup>a,b</sup>

AM1	Parameter	H	C	N	O
$f'_1(A, I)$	$\omega_A$	3.56510	2.71903	3.25328	3.61827
$f'_2(A, I)$	$\omega_A$	3.38541	2.98988	4.72260	4.95208
	$\delta_A$	0.10080	0.12996	0.32275	0.26524
MNDO	Parameter	H	C	N	O
$f'_1(A, I)$	$\omega_A$	3.44301	2.89588	3.34883	3.67920
$f'_2(A, I)$	$\omega_A$	3.48166	3.54244	5.84706	5.51133
	$\delta_A$	0.14036	0.22942	0.42741	0.31758

<sup>a</sup>For the definitions of  $f'_1(A, I)$  and  $f'_2(A, I)$ , see eqs. (10a) and (10b), respectively.

<sup>b</sup> $\omega_A$  in  $\text{\AA}^{-1}$ ,  $\delta_A$  in  $\text{\AA}$ .

**TABLE III.**  
**Parameterization of the Electrostatic Potential:**  
**Results for AM1 Wave Functions.**

Function <sup>a</sup>	$f'_1(A, I)$	$f'_2(A, I)$	$f'_2(A, I)$	—
$n_{\text{par}}$ <sup>b</sup>	1	2	2	—
Option <sup>c</sup>	$P + F$	$P$	FW <sup>d</sup>	Deorthog. <sup>e</sup>
Potential:				
rms <sup>f</sup>				
$f_{\text{vdW}} = 0.6 \cdots 1.2$	12.3	8.6	75.8	23.0
$f_{\text{vdW}} = 1.4 \cdots 2.0$	2.7	2.6	2.5	3.9
$f_{\text{vdW}} = 3.0$	0.8	0.8	0.8	1.4
rrms <sup>f</sup>				
$f_{\text{vdW}} = 0.6 \cdots 1.2$	25	18	155	47
$f_{\text{vdW}} = 1.4 \cdots 2.0$	26	25	24	38
$f_{\text{vdW}} = 3.0$	22	22	22	36
Field components:				
rms <sup>f</sup>				
$f_{\text{vdW}} = 0.6 \cdots 1.2$	31.6	55.9	423.4	39.0
$f_{\text{vdW}} = 1.4 \cdots 2.0$	2.0	2.1	2.4	1.9
$f_{\text{vdW}} = 3.0$	0.4	0.4	0.4	0.3
rrms <sup>f</sup>				
$f_{\text{vdW}} = 0.6 \cdots 1.2$	54	96	727	67
$f_{\text{vdW}} = 1.4 \cdots 2.0$	58	63	70	56
$f_{\text{vdW}} = 3.0$	60	60	61	44

<sup>a</sup>For the definition of  $f'_1(A, I)$  and  $f'_2(A, I)$ , see eqs. (10a) and (10b), respectively.

<sup>b</sup>Number of parameters.

<sup>c</sup>In the parameterizations, only potentials ( $P$ ) or potentials and field components ( $P + F$ ) are considered. The number of reference functions is 1771 ( $P$ ,  $f_{\text{vdW}} = 0.6, 0.8, 1.0, 1.2$ ) and 996 ( $F$ ,  $f_{\text{vdW}} = 0.8$ ), respectively.

<sup>d</sup>Parameterization of Ford and Wang; see ref. 35.

<sup>e</sup>Potentials and field components have been computed *ab initio* after deorthogonalization of the semiempirical wave function in a STO-6G basis.

<sup>f</sup>The rms and rrms errors refer to the original data set, which contains 4216, 10,651, and 6751 reference functions (RHF / 6-31G\* data) for the electrostatic potential on the molecular surfaces  $f_{\text{vdW}} = 0.6 \cdots 1.2$ ,  $f_{\text{vdW}} = 1.4 \cdots 2.0$ , and  $f_{\text{vdW}} = 3.0$ , respectively. The number in reference functions for the electric field is three times as large. The rms values are reported in kcal/mol e (potentials) and kcal/mol e Å (field components), respectively; rrms values are in percent.

ally underestimated. The mean deviation of the electric field is nearly zero, but this only reflects a statistically even distribution of negative and positive field strengths. The mean deviation of absolute values, however, is always positive and nearly as large as the mean absolute deviation. Thus the electric field is nearly always overestimated. This systematic error has also been observed in all the other parameterizations not reported here and resisted any attempt to be eliminated.

A closer examination of the statistical results does not reveal any specific weaknesses of our method for certain classes of molecules or bonding

**TABLE IV.**  
**Parameterization of the Electrostatic Potential:**  
**Results for MNDO Wave Functions.**

Function <sup>a</sup>	$f'_1(A, I)$	$f'_2(A, I)$	—
$n_{\text{par}}$ <sup>b</sup>	1	2	—
Option <sup>c</sup>	$P + F$	$P$	Deorthog. <sup>d</sup>
Potential:			
rms <sup>e</sup>			
$f_{\text{vdW}} = 0.6 \cdots 1.2$	13.7	9.9	22.4
$f_{\text{vdW}} = 1.4 \cdots 2.0$	3.3	3.1	3.2
$f_{\text{vdW}} = 3.0$	1.0	1.0	1.1
rrms <sup>e</sup>			
$f_{\text{vdW}} = 0.6 \cdots 1.2$	28	20	46
$f_{\text{vdW}} = 1.4 \cdots 2.0$	31	30	31
$f_{\text{vdW}} = 3.0$	25	25	30
Field components:			
rms <sup>f</sup>			
$f_{\text{vdW}} = 0.6 \cdots 1.2$	27.1	60.4	35.1
$f_{\text{vdW}} = 1.4 \cdots 2.0$	2.1	2.2	1.4
$f_{\text{vdW}} = 3.0$	0.4	0.4	0.3
rrms <sup>f</sup>			
$f_{\text{vdW}} = 0.6 \cdots 1.2$	47	104	60
$f_{\text{vdW}} = 1.4 \cdots 2.0$	63	64	41
$f_{\text{vdW}} = 3.0$	58	58	40

<sup>a-c</sup>See footnotes a–c in Table III.

<sup>d-e</sup>See footnotes e–f in Table III.

situations. In regions near the molecules, the relative errors vary between 17 and 34% for the electrostatic potential and between 41 and 64% for the electric field, respectively. Satisfactorily, the errors for the molecules not considered in the parameterization process are of about the same magnitude as for the others. More details may be found elsewhere.<sup>20</sup>

The present parameterization seems to predict electrostatic potentials and electric fields which, in an overall sense, are superior to those from the original parameterization by Ford and Wang.<sup>35</sup> However, the latter is generally more accurate for calculation of electrostatic potential minima. Notable exceptions are CO<sub>2</sub>, for which the Ford and Wang parameterization predicts the minimum to be too deep, and H<sub>2</sub>O, for which it incorrectly predicts the minimum to lie on the C<sub>2</sub> axis (see Table V). The data in Table III, however, reveal severe weaknesses in their parameterization as far as the present application is concerned. To place the electrostatic potential minima at the greater distances from the nuclei given by the *ab initio* calculations, Ford and Wang had to make their function  $f'_2(A, I)$  too steep for  $r_{AI} \rightarrow 0$ . This is clearly illustrated in Figures 1 and 2, which show the rrms errors of potentials and field strengths as

**TABLE V.**  
**Parameterization of the Electrostatic Potential: Potential Minima (AM1 Results).<sup>a</sup>**

Function <sup>b</sup>	$f'_1(A, I)$	$f'_2(A, I)$	$f'_2(A, I)$	—	—
$n_{\text{par}}^c$	1	2	2	—	—
Option <sup>d</sup>	$P + F$	$P$	FW <sup>e</sup>	Deorthog. <sup>f</sup>	<i>Ab initio</i> <sup>g</sup>
Distances <sup>h,i,j</sup>					
H <sub>2</sub> O	1.09	1.14	1.21 <sup>k</sup>	0.84	1.19
NH <sub>3</sub>	0.92	1.04	1.22	0.89	1.21
CH <sub>2</sub> O	1.16	1.17	1.23	0.87	1.23
CO <sub>2</sub>	1.37	1.30	1.28	1.01	1.43
C <sub>6</sub> H <sub>6</sub>	1.75	1.66	1.56	1.23	1.77 <sup>l</sup>
C <sub>2</sub> H <sub>4</sub>	1.90	1.86	1.87	1.23	1.53
Potentials <sup>h,i,j</sup>					
H <sub>2</sub> O	-52.4	-56.6	-58.8 <sup>k</sup>	-73.2	-63.2
NH <sub>3</sub>	-104.9	-98.3	-83.4	-112.7	-87.9
CH <sub>2</sub> O	-43.1	-48.4	-52.4	-64.4	-48.5
CO <sub>2</sub>	-19.9	-24.3	-28.9	-14.6	-18.3
C <sub>6</sub> H <sub>6</sub>	-15.0	-16.7	-20.5	-41.0	-20.0 <sup>l</sup>
C <sub>2</sub> H <sub>4</sub>	-7.1	-7.8	-9.0	-33.5	-24.7

<sup>a</sup>Distances in Å, potentials in kcal/mol e.<sup>b-e</sup>See footnotes a-d in Table III.<sup>f</sup>Potential minima have been computed *ab initio* after deorthogonalization of the AM1 wave function in a STO-4G basis. See ref. 55.<sup>g</sup>RHF/6-31G\* reference data, this work.<sup>h</sup>Distance between the potential minimum and the oxygen atom (H<sub>2</sub>O, CH<sub>2</sub>O, CO<sub>2</sub>), the nitrogen atom (NH<sub>3</sub>), or the center of inversion (C<sub>6</sub>H<sub>6</sub>, C<sub>2</sub>H<sub>4</sub>).<sup>i</sup>The potential minimum is located in the direction of the lone pairs (H<sub>2</sub>O, CH<sub>2</sub>O) or of the highest rotational axis (CO<sub>2</sub>, NH<sub>3</sub>, C<sub>6</sub>H<sub>6</sub>, C<sub>2</sub>H<sub>4</sub> [perpendicular to the molecular plane]), if not stated otherwise. No details about the location of potential minima with deorthogonalized wave functions are given in ref. 55.<sup>j</sup>Molecular geometries and potential minima have been optimized using the same quantum chemical method.<sup>k</sup>The potential minimum is located in the C<sub>2</sub> axis.<sup>l</sup>The field gradient has two positive eigenvalues (i.e., the stationary point is a second-order saddle point). The potential minimum is located outside the C<sub>6</sub> axis, 1.722 Å above the molecular plane and 1.853 Å away from the center of inversion (-20.7 kcal/mol e).

functions of  $f_{\text{vdW}}$ . The weaknesses of the variant suggested by Ford and Wang prevent their parameterization from being used in a coupled QM/MM scheme since the balance between attractive and repulsive forces can no longer be guaranteed for small distances between QM and MM fragments or molecules, and thus the danger of a collapse occurring in a QM/MM geometry optimization cannot be ruled out.

In the asymptotic region of the functions  $f'_1(A, I)$  and  $f'_2(A, I)$ —that is, for  $f_{\text{vdW}} > \sim 1.5$  (which translates into a distance of about 2 to 3 Å from the molecule)—the rrms deviations amount to 20% (electrostatic potential) and 60% (electric field strength) for AM1, respectively, and are slightly larger for MNDO (25% and 60%, respectively). These values are independent of the choice for  $f'(A, I)$  and the specific parameterization protocol. The performance of the different semiempirical models to compute electrostatic potentials and related quantities in the asymptotic region thus only reflects the quality of the semiempirical wave

function but is not correlated with the quality of the specific parameterization. Important examples are the frequently discussed potential maps and potential derived (PD) charges. Both applications have been investigated by Ford and Wang thoroughly<sup>35,59</sup> and will not be examined here.

Since the early work on CNDO/2 electrostatic potentials by Giessner-Prettre and Pullman,<sup>28</sup> the quasi *ab initio* approach based on deorthogonalized wave functions has found wide appreciation in semiempirical methodology. Although there is no lack of more efficient and thus less computer demanding purely semiempirical schemes in the literature,<sup>30-32</sup> our work and that of Ford and Wang probably present the first attempt to correct for the errors of semiempirical potentials by a specific parameterization. The statistical results (Tables III and IV) demonstrate the success of the parameterization and even show a considerable improvement over the hitherto generally preferred quasi *ab initio* approach. Although the use of deorthogonalized wave functions leads to slightly



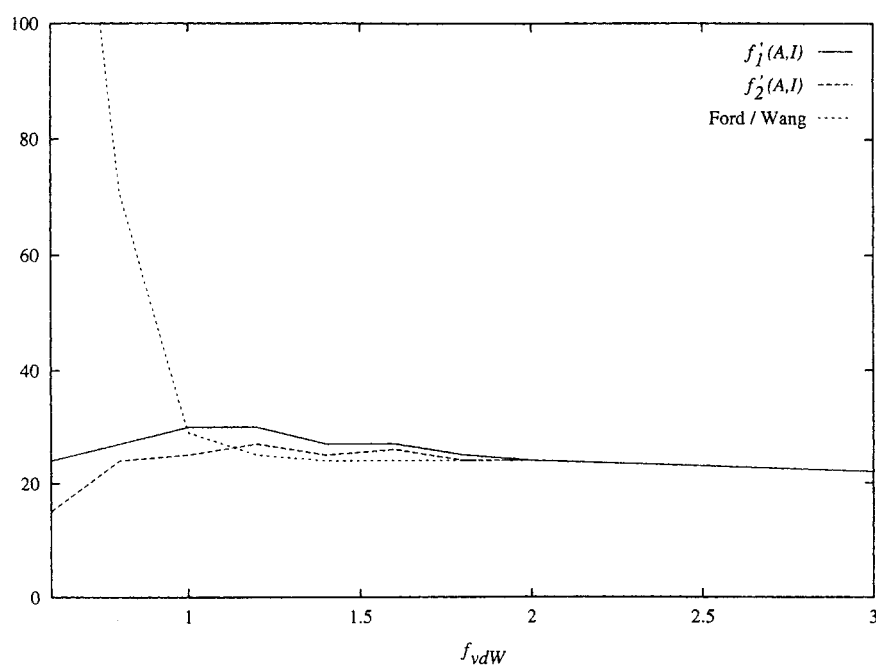
**TABLE VI.**  
**Parameterization of the Electrostatic Potential: Potential Minima (MNDO Results).<sup>a</sup>**

Function <sup>b</sup>	$f'_1(A, I)$	$f'_1(A, I)$	—	—
$n_{\text{par}}^c$	1	2	—	—
Option <sup>d</sup>	$P + F$	$P$	Deorthog. <sup>e</sup>	<i>Ab initio</i> <sup>f</sup>
Distances <sup>g, h, i</sup>				
H <sub>2</sub> O	1.06	1.10	0.86	1.19 <sup>j</sup>
NH <sub>3</sub>	0.91	1.06	0.89	1.21
CH <sub>2</sub> O	1.14	1.14	0.84 <sup>j</sup>	1.23 <sup>j</sup>
CO <sub>2</sub>	1.24	1.19	0.97	1.43
C <sub>6</sub> H <sub>6</sub>	1.93	1.65	1.33	1.77 <sup>k</sup>
C <sub>2</sub> H <sub>4</sub>	2.08	1.88	1.26	1.53
Potentials <sup>g, h, i</sup>				
H <sub>2</sub> O	−57.1	−63.2	−86.5	−63.2 <sup>j</sup>
NH <sub>3</sub>	−92.6	−91.3	−105.1	−87.9
CH <sub>2</sub> O	−49.2	−56.5	−76.3 <sup>j</sup>	−48.5 <sup>j</sup>
CO <sub>2</sub>	−29.0	−36.0	−26.5	−18.3
C <sub>6</sub> H <sub>6</sub>	−6.8	−9.0	−25.5	−20.0 <sup>k</sup>
C <sub>2</sub> H <sub>4</sub>	−2.9	−4.1	−24.9	−24.7

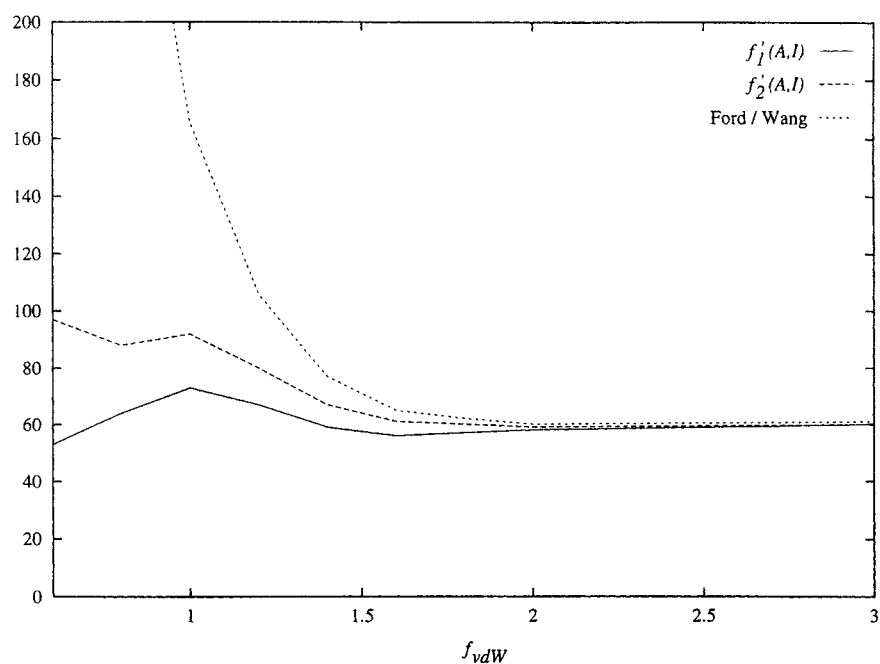
<sup>a, f, g, i, k</sup> See footnotes a, g, h, j, l in Table V.<sup>b, d</sup> See footnotes a–c in Table III.<sup>e</sup> Potential minima have been computed *ab initio* after deorthogonalization of the MNDO wave function in a STO-4G basis. See ref. 54.<sup>h</sup> The potential minima are located in the direction of the highest rotational axis (C<sub>2</sub>H<sub>4</sub>: perpendicular to the molecular plane), if not stated otherwise. There are no full details in ref. 54 about the location of potential minima computed with a deorthogonalized MNDO wave function.<sup>i</sup> The potential minimum is located in the direction of the oxygen lone pair.**TABLE VII.**  
**Statistical Analysis of Semiempirical Potentials and Field Strengths.<sup>a, b</sup>**

AM1		Potential, $f_{\text{vdW}} =$			Field Strength, $f_{\text{vdW}} =$		
		0.6 ... 1.2	1.4 ... 2.0	3.0	0.6 ... 1.2	1.4 ... 2.0	3.0
$f'_1(A, I)$	$\Delta_{\pm}$	−4.08	−0.27	−0.08	−0.27	−0.01	0.00
	$\Delta_{\text{abs}}$	8.35	2.04	0.06	14.06	1.30	0.27
	$\Delta_{\pm}$	−5.95	−1.58	−0.49	12.00	1.10	0.26
$f'_2(A, I)$	$\Delta_{\pm}$	−0.67	−0.32	−0.08	−0.30	−0.01	0.00
	$\Delta_{\text{abs}}$	5.79	1.97	0.60	23.65	1.42	0.27
	$\Delta_{\pm}$	−1.65	−1.50	−0.49	22.42	1.27	0.26
MNDO		0.6 ... 1.2	1.4 ... 2.0	3.0	0.6 ... 1.2	1.4 ... 2.0	3.0
$f'_1(A, I)$	$\Delta_{\pm}$	−3.67	−0.05	−0.06	−0.16	−0.01	0.00
	$\Delta_{\text{abs}}$	9.49	2.50	0.73	13.55	1.42	0.26
	$\Delta_{\text{abs}}$	−5.93	−1.92	−0.61	10.72	0.95	0.21
$f'_2(A, I)$	$\Delta_{\pm}$	−1.04	−0.16	−0.06	−0.25	−0.01	0.00
	$\Delta_{\text{abs}}$	6.94	2.42	0.73	24.78	1.42	0.26
	$\Delta_{\text{abs}}$	−2.00	−1.87	−0.61	23.16	1.05	0.21

<sup>a</sup> Electrostatic potentials in kcal/mol e, electric field strengths in kcal/mol Å e.<sup>b</sup> For a definition of the statistical quantities, see eq. (9); they refer to RHF/6-31G\* data. The functions  $f'_1(A, I)$  and  $f'_2(A, I)$  are defined in eqs. (10a) and (10b), respectively.



**FIGURE 1.** The rms error of the AM1 electrostatic potential as a function of  $f_{vdW}$ . Each curve comprises 21,618 data computed for the reference molecules.



**FIGURE 2.** The rms error of the AM1 electric field strength as a function of  $f_{vdW}$ . Each curve comprises 64,854 data computed for the reference molecules.

better electric fields, it fails to compute the minima of electrostatic potentials accurately. Quasi *ab initio* approaches tend to overestimate the absolute values of these minima and to predict them much too close to the molecular periphery (see Tables IV and V). In this respect, the purely semiempirical approaches are much more accurate, as is demonstrated by the data in Tables IV and V and the potential curves for the oxirane molecule, which are shown for illustration (Figures 3 through 5).

## Partial Charges

### GENERAL CONSIDERATIONS

The proper definition of partial charges for the MM fragment is one of the most important and most difficult steps in the development of QM/MM models. The main problem arises from the fact that partial charges are not observable quantities (i.e. they do not correspond to the expectation value of a quantum mechanical operator). Thus, any possible definition is arbitrary.

As noted in the previous section, the Coulomb interaction between two sets of point charges may be expressed as the energy of either set of point charges in the electrostatic potential of the other.

Molecules interact via their continuous charge distribution, generating electrostatic potentials which may only be approximated by discrete sets of point charges. According to eq. (1), the Coulomb interaction energy in the QM/MM model is evaluated from the quantum chemical electrostatic potential arising from a continuous charge distribution in the QM part and the classical partial charges in the MM part. For a realistic description of this interaction, the partial charges should thus be chosen such that their electrostatic potential

$$\langle \Phi^A \rangle \approx \Phi_{\text{model}}^A = \sum_I \frac{q_I}{r_{AI}} \quad (11)$$

reproduces the "real" potential arising from the continuous charge distribution in the MM part. This motivates the use of potential derived atomic point charges.

The direct use of PD charges in a combined QM/MM scheme, however, is not feasible since it requires a quantum chemical calculation for the classical region of the supermolecule. Instead we need an empirical model which allows us to predict PD charges with an acceptable accuracy. In 1952, Sanderson proposed the principle of electronegativity equalization, which states that the electronegativities of atoms (or, more generally, of

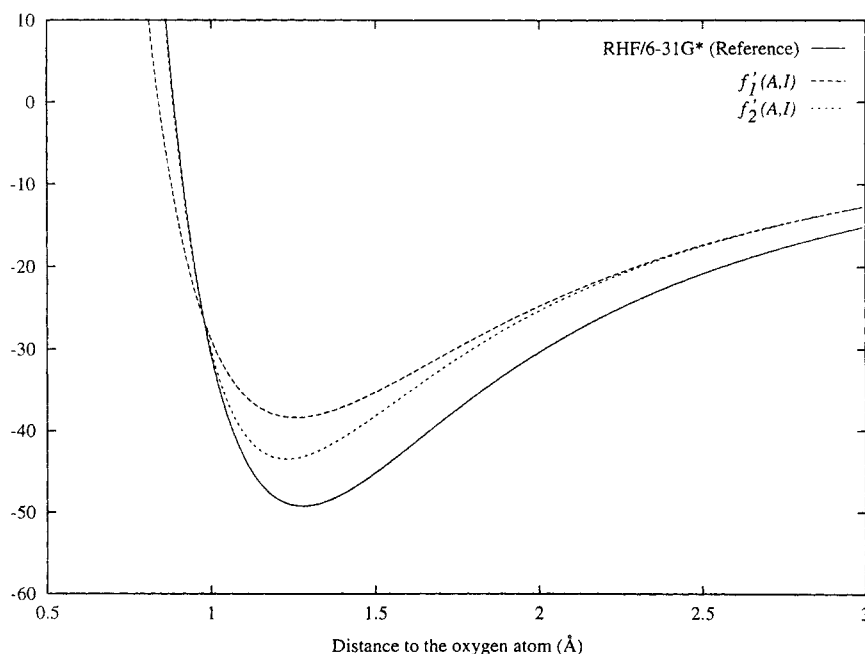
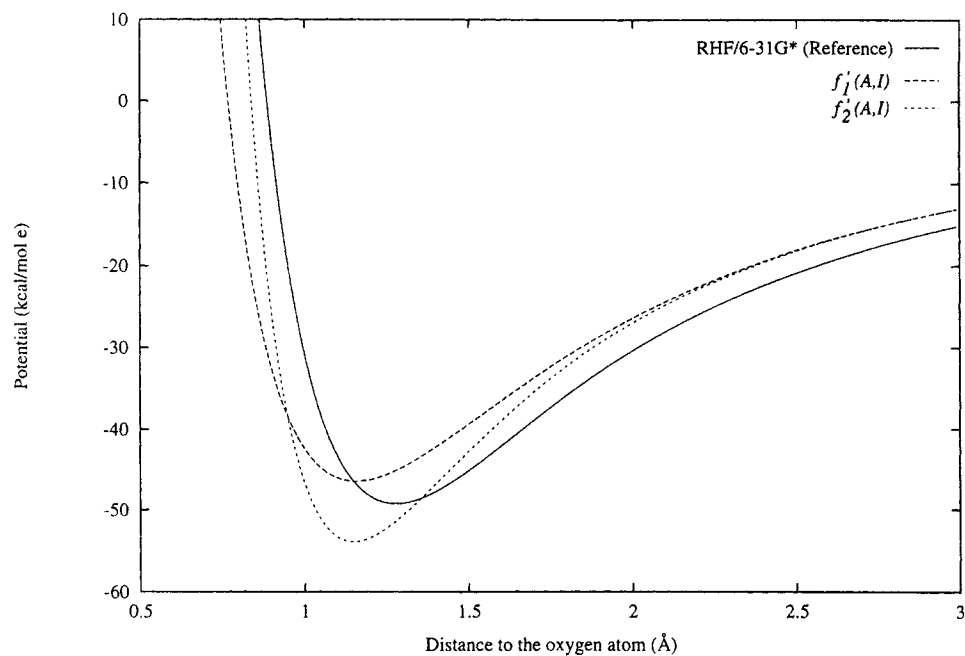
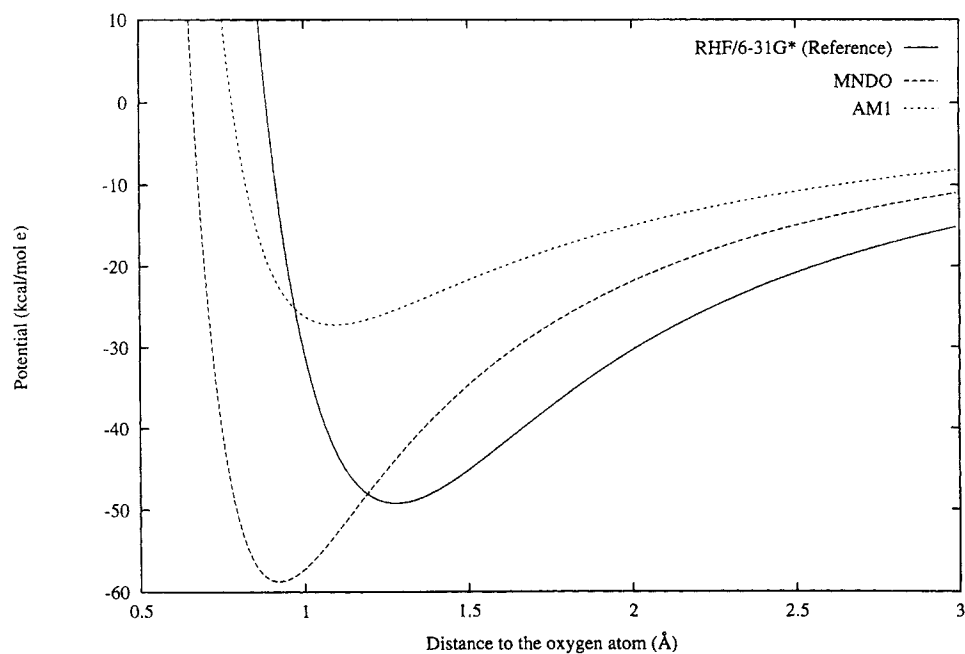


FIGURE 3. AM1 electrostatic potential in the  $C_2$  axis of the oxirane molecule.



**FIGURE 4.** MNDO electrostatic potential in the  $C_2$  axis of the oxirane molecule.



**FIGURE 5.** Electrostatic potential in the  $C_2$  axis of the oxirane molecule computed with deorthogonalized wave functions.

atomic orbitals) equalize during the process of molecule formation.<sup>37-39</sup> Parr and co-workers proved this hypothesis in 1978 on the basis of Hohenberg-Kohn density functional theory.<sup>40</sup> Further supporting arguments stem from heuristic concepts which have been considered by Politzer and Weinstein.<sup>60</sup>

Mulliken's definition<sup>61</sup> of the atomic electronegativity  $\chi_I^0$  as the arithmetic mean of its ionization potential  $IP_I$  and its electron affinity  $EA_I$  directly connects electronegativity and charge state and thus allows us to calculate partial charges from a reasonable model for the total energy of the molecule. Several charge models are based on this idea (e.g., those proposed by Gasteiger and Marsili,<sup>41</sup> Mortier et al.,<sup>42</sup> and Rappé and Goddard<sup>43</sup>). We follow the charge equilibration model of Rappé and Goddard but redevelop it in a semiempirical formulation consistent with the other components of our QM/MM schemes.

The total energy of a molecule or a molecular fragment  $X$  is written as the sum of all one-center and two-center contributions:

$$E(X) = \sum_I E_I + \sum_{I < J} \sum_I E_{IJ} \quad (12)$$

If one assumes that the atomic energy  $E_I$  is a continuous, differentiable function of the atomic charge  $q_I$ , it may be developed as a Taylor series around  $q_I = 0$  which is terminated after the quadratic term:

$$E_I = E_I^0 + q_I \left( \frac{\partial E}{\partial q_I} \right)_0 + \frac{1}{2} q_I^2 \left( \frac{\partial^2 E}{\partial q_I^2} \right)_0 \quad (13)$$

Considering eq. (13) as an acceptable approximation for the energies of positively ( $E_I^{+1}$ ) and negatively ( $E_I^{-1}$ ) charged ions, the ionizations potential and the electron affinity may be expressed via partial derivatives of the total energy:

$$IP_I = -E_I^0 + E_I^{+1} = \left( \frac{\partial E}{\partial q_I} \right)_0 + \frac{1}{2} \left( \frac{\partial^2 E}{\partial q_I^2} \right)_0 \quad (14)$$

$$EA_I = +E_I^0 - E_I^{-1} = \left( \frac{\partial E}{\partial q_I} \right)_0 - \frac{1}{2} \left( \frac{\partial^2 E}{\partial q_I^2} \right)_0 \quad (15)$$

The Mulliken definition of the atomic electronegativity then yields the well-known relation<sup>40,62</sup>

$$\chi_I^0 := \frac{1}{2} (IP_I + EA_I) = \left( \frac{\partial E}{\partial q_I} \right)_0 \quad (16)$$

The difference  $J_{II}$  of the ionization potential and the electron affinity is similarly given by the second derivative of the energy:

$$J_{II} := IP_I - EA_I = \left( \frac{\partial^2 E}{\partial q_I^2} \right)_0 \quad (17)$$

The two-center terms  $E_{IJ}$  in eq. (12) are treated as simple Coulomb interactions:

$$E_{IJ} = J_{IJ} q_I q_J \quad (18)$$

By comparison of eqs. (12), (13), and (16)–(18), one finds that

$$E(X) = \sum_I \left( E_I^0 + \chi_I^0 q_I + \frac{1}{2} J_{II} q_I^2 \right) + \sum_{I < J} \sum_I J_{IJ} q_I q_J \quad (19)$$

In analogy to eq. (16), the electronegativity  $\chi_I$  of an atom  $I$  within a molecule may be expressed as the first derivative of the total energy with respect to the charge  $q_I$ . Differentiation of eq. (19) yields

$$\chi_I = \chi_I^0 + J_{II} q_I + \sum_{J \neq I} J_{IJ} q_J = \chi_I^0 + \sum_J J_{IJ} q_J \quad (20)$$

Application of the principle of electronegativity equalization ( $\chi_1 = \chi_2 = \dots = \chi_N$ ) leads to a system of  $N - 1$  linear equations with  $N$  unknowns ( $q_I$ ):

$$\chi_1^0 - \chi_I^0 = \sum_J (J_{IJ} - J_{1J}) q_J \quad (I = 2, \dots, N) \quad (21)$$

Imposing the constraint that the sum of the partial charges must equal the (known) total charge  $Q^X$ , one obtains a uniquely solvable system of equations which in matrix notation reads

$$\mathbf{D} = \mathbf{C} \mathbf{q} \quad (22)$$

The vector  $\mathbf{q}$  contains all the partial charges ( $q_1, q_2, \dots, q_N$ ), whereas the components of  $\mathbf{D}$  and the elements of  $\mathbf{C}$  are defined as follows:

$$D_I := \chi_1^0 - \chi_I^0 + \delta_{1,I} Q^X \quad (23)$$

$$C_{IJ} := J_{IJ} - J_{1J} + \delta_{1,I} J_{1J} \quad (24)$$

The partial charges may now easily be obtained by inversion of  $\mathbf{C}$ .

$$\mathbf{q} = \mathbf{C}^{-1} \mathbf{D} \quad (25)$$

After discussing the general scheme, we have to find suitable expressions for the one- and two-

center terms  $J_{II}$  and  $J_{IJ}$ . For the two-center terms, the assumption of a totally classical interaction  $J_{IJ} = V^{IJ}$  would be consistent with the semiempirical formulas we have proposed for the electrostatic potential. Such a treatment neglects, however, overlap effects for small atomic distances (e.g., in a chemical bond) and consequently leads to unrealistic results, as the work of Rappé and Goddard<sup>43</sup> and our own test calculations have shown. Thus the damping effects occurring for small distances have to be taken into account through a suitably modified effective Coulomb potential, which smoothly converges to the classical Coulomb potential in the long range. The Klopman-Ohno approximation, discussed in the last section, fulfils this requirement. For the interaction of two (*ss*) charge distributions, it reads

$$J_{IJ} = e^2 \left( r_{IJ}^2 + (\rho_0^I + \rho_0^J)^2 \right)^{-\frac{1}{2}} \quad (26)$$

The one-center terms may—at least in principle—be derived explicitly from eq. (17). To be in line with the simple approximation for the interaction of two different atoms, however, we assume that any of the one-center terms  $J_{II}$  describes an effective Coulomb interaction which may be represented by a Coulomb integral over *s* functions. We take  $J_{II}$  as a free parameter but impose the constraint that it is the one-center limit of the respective two-center integral:

$$\rho_0^I = \frac{e^2}{2J_{II}} \quad (27)$$

## PARAMETERIZATIONS

The error function to be optimized with respect to the parameters  $\chi_i^0$  and  $J_{II}$  is defined as a sum over all  $n_{\text{ref}}^q$  squared deviations between the model charges  $q^n$  and the *ab initio* reference charges  $q_{\text{ref}}^n$ .

$$\text{SSQ} = \sum_{n=1}^{n_{\text{ref}}^q} (q^n - q_{\text{ref}}^n)^2 \quad (28)$$

Although the constraint of yielding the correct molecular charge introduces linear dependencies between the partial charges, this does not cause real problems since both the QEq model and the *ab initio* reference data obey the same constraint.

A first set of reference data consisted of RHF/6-31G\*/PD charges obtained by linear optimization of the model potential  $\Phi_{\text{model}}^A$  [see eq. (11)] accord-

ing to the protocol described by Besler et al.<sup>51</sup> In complete analogy to the parameterizations discussed in the last section, the Connolly algorithm was applied to generate points with a density of  $1.0 \text{ \AA}^{-2}$  on four van der Waals surfaces ( $f_{\text{vdW}} = 1.4, 1.6, 1.8, 2.0$ ) of each molecule in the data set. RHF/6-31G\* Mulliken charges<sup>47,63</sup> served as alternative reference functions. We admit that the use of charges derived from population analyses is not motivated by our theoretical considerations and that the Mulliken population analysis in particular has been criticized repeatedly in the literature<sup>31,36,64,65</sup> due to its marked basis set dependence and the hardly justifiable symmetrical separation of heteronuclear bond populations. Nonetheless, it seemed interesting to check which set of reference charges the QEq model does more justice to. The charge models based on these two parameterizations are denoted by QEq/PD and QEq/Mulliken, respectively.

The reference data set for the parameterizations originally included all the molecules listed in Table I, with the only exception of ethane, whose weak electrostatic potential could not be reproduced satisfactorily by a simple point charge model with only one charge per atom. This failure of point charge models is not limited to ethane but occurs for other *n*-alkanes as well, as has been pointed out by Williams.<sup>66</sup> Even though there were no problems with the Mulliken charges of ethane, we decided to be consistent and exclude ethane from all parameterizations, but include it in the statistical analysis. The remaining 23 molecules provided a total of 135 reference functions, which could easily be handled by our parameterization program without any need for a further data reduction.

The molecules nitromethane (11) and carbon monoxide (13) turned out to be problematic in the parameterization. The QEq model always dramatically overestimated the charges for CO, irrespective of specific changes in the parameterization protocol, whereas the charge distribution of the nitro group was successfully reproduced only at the expense of physically questionable parameters ( $J_{NN} < J_{CC}$ ). Both the MNDO integrals ( $s_I s_I, s_I s_I$ ) and the explicitly calculated quantities  $J_{II}$  [starting from eq. (17) and applying the MNDO approximation] show the opposite trend. Obviously, the QEq model does not adequately treat molecules which are characterized by valence structures with separated charges ( $^+ \text{NO}_2^-$ ,  $^- \text{CO}^+$ ). This may be traced back to the crude model assumption that a given element can be described by one mean valence

state, irrespective of its true chemical environment. It is probably necessary to define separate parameters  $\chi_K^0$  and  $J_{KK}$  for structures like these. Instead of introducing such refinements, we decided to exclude nitromethane and carbon monoxide and to work with the remaining 21 reference molecules.

The parameterization of the QEq/PD model proceeded smoothly and yielded acceptable values for both  $\chi_I^0$  and  $J_{II}$  (see Table VIII). In the case of the QEq/Mulliken model, the results were fairly insensitive to rather large parameter changes, and the optimizations tended to give unrealistic values for  $J_{HH}$  ( $> 20$  eV, too large) and  $\chi_H^0$  ( $< 3$  eV, too low). Hence we decided to introduce three constraints derived from the QEq/PD parameterizations (M denotes QEq/Mulliken, PD denotes QEq/PD):

$$J_{HH}^M = J_{HH}^{PD} \quad (29)$$

$$J_{II}^M = J_{CC}^M + (J_{II}^{PD} - J_{CC}^{PD}) \quad (I = N, O)$$

These additional constraints led to more realistic values for  $\chi_I^0$  and  $J_{II}$  without seriously affecting the quality of the statistical results.

Table VIII not only compares our final parameters with the values in the original QEq model proposed by Rappé and Goddard but also shows the values of  $\chi_I^0$  and  $J_{II}$ , which are obtained from experimental ionization potentials and electron affinities via eqs. (16) and (17). The latter are equal to twice the absolute hardness  $\eta_I$ , as defined by Parr and Pearson.<sup>62</sup> The most significant differences between the model parameters and the experimental values are seen in the sequences  $\chi_H^0 \leftrightarrow \chi_C^0$  and  $J_{NN} \leftrightarrow J_{OO}$ . However, the experimental

values refer to free atoms while the model parameters characterize atoms in a molecule. The experimental atomic data lead to the counterintuitive conclusion that hydrogen should be more electronegative than carbon. Rappé and Goddard reason that the effective electron affinity of the hydrogen atom in a molecule should be smaller, thus causing a lower electronegativity as well.<sup>43</sup> Other electronegativity scales based on molecular rather than atomic data also predict hydrogen atoms to be less electronegative than carbon atoms (e.g., Pauling's scale<sup>67</sup>).

## RESULTS AND DISCUSSION

The correlations between QEq/PD and 6-31G\*/PD charges and those between the QEq/Mulliken and 6-31G\*/Mulliken charges for all the reference molecules, including ethane, are shown in Figures 6 and 7. The variation around the straight line characterizing ideal correlation (unit slope) is larger for the PD charges. This is confirmed by the statistical analysis [ $y(\text{QEq}) = a \cdot x(6\text{-}31\text{G}^*)$ ], which yields correlation coefficients of 0.959 and 0.991 and slopes of 0.970 and 1.002 for PD and Mulliken charges, respectively. The rms deviations amount to 0.105 e and 0.050 e for PD and Mulliken charges, respectively. Apparently, the basic assumptions of the QEq model cause charge transfers which mirror the electron distribution in a way reminiscent of the Mulliken population analysis.

A closer inspection of the data shows that the opposite CH polarization of methyl groups directly bound to electronegative atoms is the main

**TABLE VIII.**  
Parameters of the QEq Model.<sup>a</sup>

		H	C	N	O
QEq / PD	$\chi_I^0$	4.42211	5.07305	7.73699	8.27885
	$J_{II}$	13.84036	10.06444	12.96908	14.93241
QEq / Mulliken	$\chi_I^0$	3.87366	5.17653	7.54758	8.91322
	$J_{II}$	13.84036	11.22625	14.13089	16.09422
Rappé / Goddard <sup>b</sup>	$\chi_I^0$	4.5280	5.343	6.899	8.741
	$J_{II}$	13.8904	10.126	11.760	13.364
Parr / Pearson <sup>c</sup>	$\chi_I^0$	7.17	6.27	7.27	7.53
	$J_{II}$	12.84	10.00	14.54	12.16

<sup>a</sup>All values in eV.

<sup>b</sup>Values reported by Rappé and Goddard.<sup>43</sup>

<sup>c</sup>From experimental atomic ionization potentials and electron affinities.<sup>62</sup>

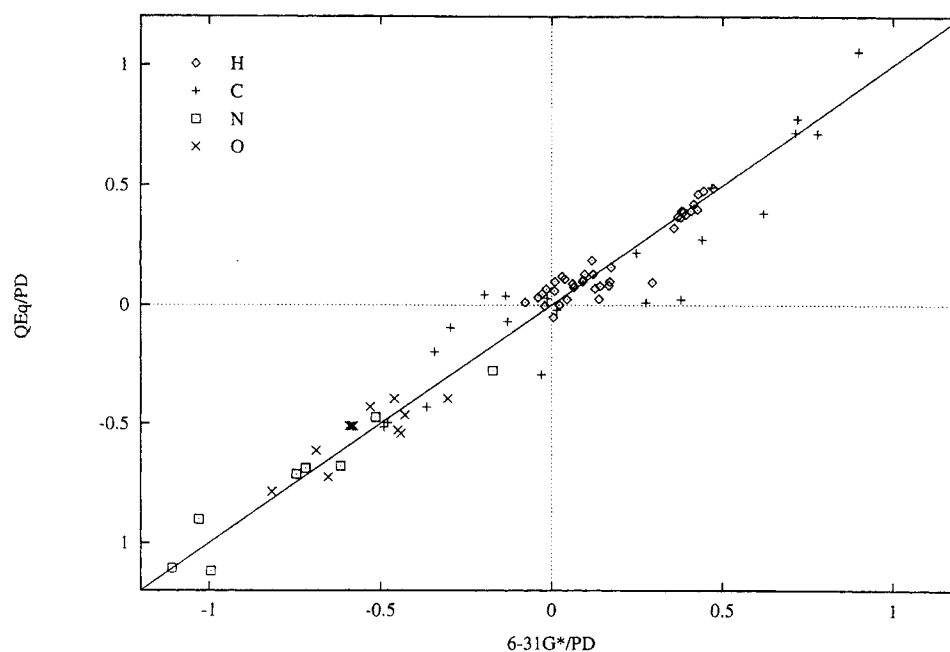


FIGURE 6. Potential derived charges of the reference molecules.

qualitative difference between PD ( $C^+H^-$ ) and Mulliken ( $C^-H^+$ ) charges. The carbon atom of methylamine, for instance, carries a large positive PD charge (+0.379 e) but a large negative Mulliken charge (−0.296 e). The QEq model only re-

produces the Mulliken charge with reasonable accuracy (Table IX). In general, the QEq/PD model shows the largest deviations for carbon atom charges, as can be seen in Figure 6. This need not be a specific weakness of the QEq/PD model since

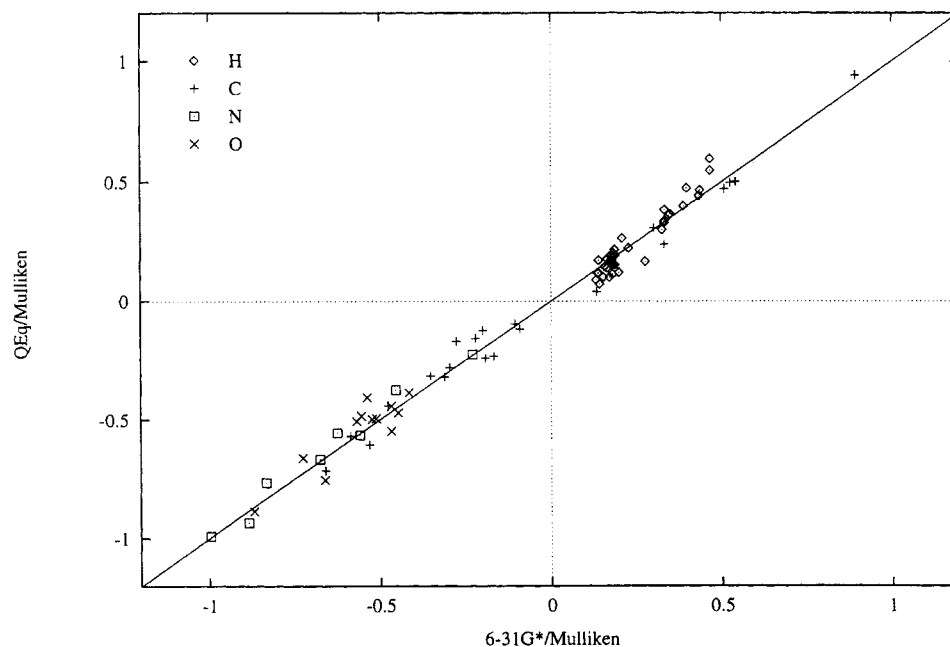


FIGURE 7. Mulliken charges of the reference molecules.



carbon atoms usually constitute the molecular backbones and thus have the largest distances to the molecular surfaces relevant for the PD charge fits. These buried charges show the lowest statistical significance in the charge fits, as has been repeatedly pointed out in recent work.<sup>68-70</sup> The relative error of the potential fit for methanol, for instance, increases from 10.7% to only 11.5% if the fit is subject to the weighted (0.5%) constraint to reproduce Mulliken charges (see Table I in ref. 68). In this case, the charge of the carbon atom is reduced from 0.196 e to 0.022 e (for comparison, QEq/PD: 0.011 e), whereas the changes for all the other atoms are less than 0.05 e.

On the whole, the statistical results of the QEq/PD parameterization may be regarded as satisfactory. The errors are similar to those of the frequently recommended PD charges computed with deorthogonalized MNDO wave functions.<sup>51,71</sup> Comparing MNDO/PD and 6-31G\*/PD charges for a similar set of reference molecules, Orozco and Luque obtained a correlation coefficient of 0.98, a slope of 1.32, and an rms error for the unscaled charges of 0.14 e.<sup>71</sup> In a similar study, Besler et al. reported a correlation coefficient of 0.97 and a slope of 1.42.<sup>51</sup>

Some of the most important QEq/PD charges are classified in Table X according to specific bond types, a more complete list of which may be found elsewhere.<sup>20</sup> Apart from the NO bond in nitromethane, the CO bond in carbon monoxide, and the CH bonds, which have all been discussed earlier, the bond polarities are reproduced reasonably well by the QEq/PD model. In particular, one observes the same characteristic trends for single, double, and triple bonds of heteroatoms (e.g., CN) or for heteroatoms bound to different neighbors (e.g., H<sub>2</sub>O, CH<sub>3</sub>OH, H<sub>2</sub>O<sub>2</sub>) in both the QEq/PD and the 6-31G\*/PD charges. This is not at all self-evident considering that our model does not explicitly discriminate between equivalent atoms

in different chemical surroundings, in contrast to the PEOE (Partial Equalization of Orbital Electronegativity) formalism proposed by Gasteiger and Marsili.<sup>41</sup> One may conclude that the distance-dependent Coulomb interaction  $J_{IJ}$ , which has no counterpart in the PEOE formalism, accounts for the correct differentiation between various bonding situations.

The dipole moments computed from QEq/PD charges may serve for further validation (see Table XI). Despite the relatively large mean variation, the overall agreement of QEq/PD dipole moments with experimental and 6-31G\*/PD values is satisfactory. The mean absolute deviations are 0.33 D and 0.29 D compared to experimental and 6-31G\*/PD values, respectively (excluding CH<sub>3</sub>NO<sub>2</sub> and CO in both cases) and thus lie in the same range as the errors of semiempirical quantum chemical methods (AM1: 0.26 D, MNDO: 0.32 D, compared to experiment<sup>23</sup>).

Finally, the QEq/PD charges may be compared with the results which Rappé and Goddard have obtained using their original formulation of the QEq model.<sup>43</sup> Their model expresses the interaction terms  $J_{IJ}$  as *ab initio* two-center integrals ( $s_I s_J, s_I s_I$ ) and internally couples the one-center integrals  $J_{HH}$  to the partial charges  $q_H$  of the respective hydrogen atoms. All available data are included in Table X. While the polarization of CH bonds is calculated to be almost the same with both QEq variants, our model shows slight advantages for other bonding situations. In particular, our model proves to be superior for strongly polar molecules (e.g., NH<sub>3</sub>, NH<sub>2</sub>CHO, H<sub>2</sub>O). This is probably due to our specific parameterization, although Rappé and Goddard also introduced two parameters ( $J_{HH}$  and  $\chi_H^0$ ) which they calibrated against *ab initio* PD charges of, among others, H<sub>2</sub>O, NH<sub>3</sub>, and CH<sub>4</sub>.<sup>43</sup> Apart from the slightly better results, the QEq/PD model proposed here is conceptually simpler and less demanding numerically, since it does not require an iterative procedure<sup>43</sup> for computation of the partial charges.

Any semiempirical method can only be regarded as useful if it also proves to be successful for cases other than those considered in the parameterization. Since most of the important bonding situations characteristic of organic molecules have already been covered by the parameterization, our tests have focused on trends for several classes of similar molecules instead of studying a large variety of more complicated cases. The results are reported in detail elsewhere<sup>20</sup>; here we only give a short summary.

**TABLE IX.**  
Atomic Partial Charges in Methylamine.<sup>a</sup>

Atom	6-31G* / PD	6-31G* / Mulliken	QEq / PD	QEq / Mulliken
C	0.379	-0.296	0.024	-0.280
H(-C) <sup>b</sup>	-0.036	0.155	0.048	0.124
N	-1.030	-0.833	-0.900	-0.767
H(-N)	0.379	0.332	0.366	0.338

<sup>a</sup>All values in atomic units.

<sup>b</sup>Averaged value for the three hydrogen atoms.

**TABLE X.**  
**Atomic Partial Charges in Heteronuclear Bonds.<sup>a,b</sup>**

<i>I - J</i>	Molecule	6-31G* / PD		QEq / PD		Rappé / Goddard <sup>c</sup>	
		<i>q<sub>I</sub></i>	<i>q<sub>J</sub></i>	<i>q<sub>I</sub></i>	<i>q<sub>J</sub></i>	<i>q<sub>I</sub></i>	<i>q<sub>J</sub></i>
H—C	CH <sub>4</sub>	0.123	−0.490	0.129	−0.516	0.13	−0.52
	C <sub>2</sub> H <sub>6</sub>	0.010	−0.029	0.098	−0.293	0.13	−0.39
	C <sub>2</sub> H <sub>4</sub>	0.171	−0.342	0.098	−0.196	0.13	−0.26
	C <sub>2</sub> H <sub>2</sub>	0.295	−0.295	0.095	−0.095	0.11	−0.11
	CH <sub>3</sub> NH <sub>2</sub>	−0.035	0.379	0.048	0.024	—	—
	CH <sub>2</sub> NH	0.045	0.248	0.038	0.218	—	—
	CH <sub>3</sub> OH	−0.005	0.276	0.068	0.011	0.13	−0.09
	CH <sub>2</sub> O	0.009	0.441	0.060	0.274	0.10	0.24
H—N	NH <sub>3</sub>	0.370	−1.109	0.369	−1.107	0.23	−0.69
	CH <sub>3</sub> NH <sub>2</sub>	0.379	−1.030	0.366	−0.901	—	—
	NH <sub>2</sub> CHO	0.431	−0.994	0.450	−1.118	0.25	−0.61
	CH <sub>2</sub> NH	0.381	−0.719	0.394	−0.688	—	—
	N <sub>2</sub> H <sub>4</sub>	0.373	−0.746	0.356	−0.713	—	—
H—O	H <sub>2</sub> O	0.409	−0.817	0.393	−0.787	0.35	−0.70
	CH <sub>3</sub> OH	0.427	−0.688	0.400	−0.613	0.34	−0.66
	H <sub>2</sub> O <sub>2</sub>	0.429	−0.429	0.464	−0.464	—	—
C—N	CH <sub>3</sub> NH <sub>2</sub>	0.379	−1.030	0.024	−0.901	—	—
	CH <sub>3</sub> NNCH <sub>3</sub>	0.016	−0.172	−0.021	−0.276	—	—
	CH <sub>2</sub> NH	0.248	−0.719	0.218	−0.688	—	—
C≡N	CH <sub>3</sub> CN	0.468	−0.515	0.493	−0.473	0.22	−0.25
C≡O	CH <sub>3</sub> OH	0.276	−0.688	0.011	−0.613	−0.09	−0.66
C=O	CH <sub>2</sub> O	0.441	−0.459	0.274	−0.394	0.24	−0.44
	HCOOH	0.713	−0.579	0.722	−0.510	0.58	−0.44
	CO <sub>2</sub>	0.898	−0.449	1.055	−0.528	0.90	−0.45
C≡O	CO	0.062	−0.062	0.434	−0.434	—	—
N—O	CH <sub>3</sub> NO <sub>2</sub>	0.870	−0.485	−0.091	−0.156	—	—

<sup>a</sup>All values in atomic units.<sup>b</sup>Averaged values for hydrogen atoms bound to a common heavy atom.<sup>c</sup>QEq<sup>HIF</sup> model of Rappé and Goddard; ref. 43 (exp. geometries).

One interesting question concerns the ability of the QEq model to reproduce the influence of increasing alkyl substitution and increasing chain length on the charge of the central atom of alcohols (ethers) and amines. Potential derived charges of both alcohol (ether) oxygen and amine nitrogen atoms decrease on methyl substitution and slightly increase on extending the substituent chain length. Both effects are reproduced qualitatively by our QEq/PD model, but they are underestimated quantitatively. For instance, the 6-31G\*/PD charge of oxygen in dimethyl ether is lower than in methanol by 0.33 e, whereas the QEq/PD model predicts a decrease of only 0.18 e.<sup>20</sup>

The charge distribution in  $\alpha$ -D- and  $\beta$ -D-idose constitutes a more complicated test case for the QEq/PD model which has also been discussed in other studies on semiempirical charge schemes.<sup>59,72</sup> For easier reference, we apply the same geometries and numbering conventions for both molecules as

in a previous study on MNDO/PD charges.<sup>72,73</sup> The molecules are depicted in Figure 8; our results are summarized in Table XII.

Similar to other cases, the QEq/PD and the 6-31G\*/PD charges qualitatively agree in the most significant trends for equal atoms in different positions (e.g., compare, the charges of O-14, O-16, and O-19 with those of O-15 and O-17). In detail, however, one notices some differences between the predictions of both charge schemes. The 6-31G\*/PD charges for C-1 are computed to be different for the anomeric forms, whereas the QEq/PD model predicts nearly equal charges. The same holds true for C-2 and C-6. The large configurational changes are limited, however, to buried atoms which belong to the molecular skeletons. Thus their statistical significance may again be questioned.<sup>68-70</sup> Having this in mind, the performance of the QEq/PD model appears to be satisfactory, especially since the rms deviation with

**TABLE XI.**  
**Dipole Moments.<sup>a</sup>**

	Molecule	Experiment <sup>b</sup>	6-31G* / PD <sup>c</sup>	QEq / PD
1	Aziridine	1.90	1.94	1.71
6	CH <sub>2</sub> NH	1.97	2.20	1.91
7	CH <sub>2</sub> O	2.33	2.66	2.58
8	CH <sub>3</sub> CHO	2.75	2.98	2.92
9	CH <sub>3</sub> CN	3.92	4.06	3.28
10	CH <sub>3</sub> NH <sub>2</sub>	1.31	1.52	1.42
11	CH <sub>3</sub> NO <sub>2</sub>	3.46	4.02	4.43
13	CO	0.11	0.26	2.32
15	H <sub>2</sub> O <sub>2</sub>	1.57	1.95	2.11
16	HCOOCH <sub>3</sub>	1.77	1.99	1.69
17	HCOOH	1.41	1.60	0.91
18	CH <sub>3</sub> OH	1.70	1.85	1.82
19	N <sub>2</sub> H <sub>4</sub>	1.75	2.21	1.87
21	NH <sub>2</sub> CHO	3.73	4.10	3.16
22	NH <sub>3</sub>	1.47	1.97	1.97
23	Oxirane	1.89	2.33	2.60
24	H <sub>2</sub> O	1.85	2.25	2.17

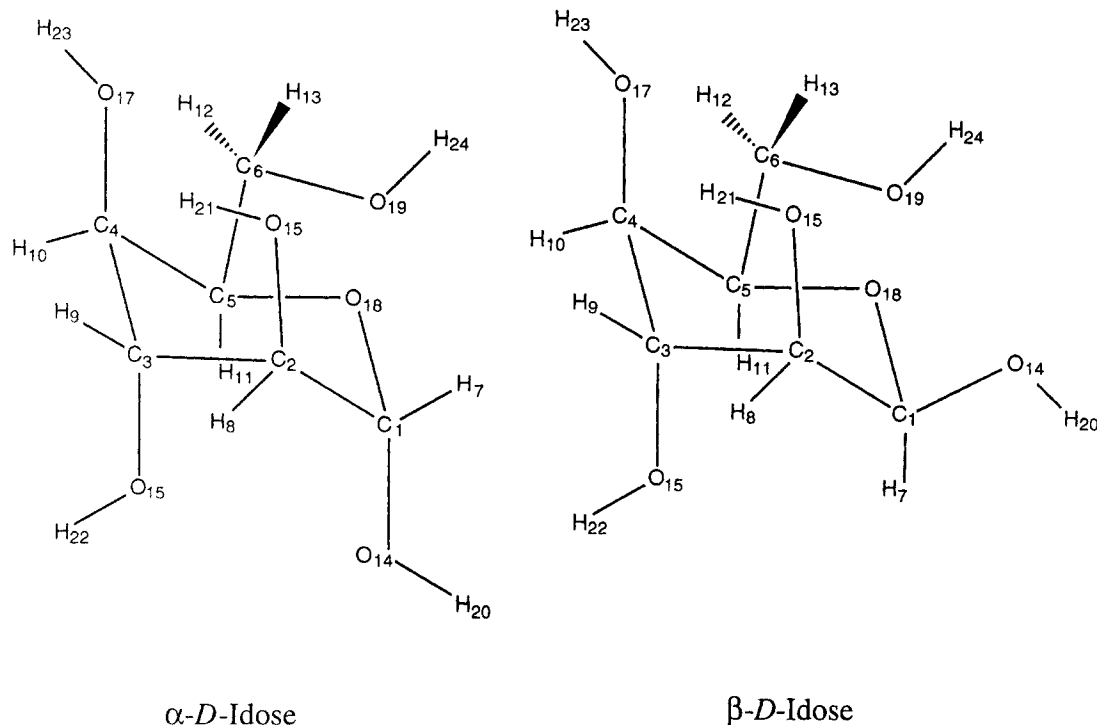
<sup>a</sup>All values in Debye (D).<sup>b</sup>From ref. 74, except aziridine [M. J. S. Dewar and W. Thiel, *J. Am. Chem. Soc.*, **99**, 4907 (1977)] and CH<sub>2</sub>NH [M. Allegrini, J. W. C. Johns, and A. R. W. McKellar, *J. Chem. Phys.*, **70**, 2829 (1979)].<sup>c</sup>Identical to RHF/6-31G\* dipole moments to within  $\pm 0.05$  D, except H<sub>2</sub>O<sub>2</sub> (1.85 D).

respect to 6-31G\*/PD charges is even lower than for the molecules considered in the parameterization (0.079 e vs. 0.105 e).

## Conclusions

Efficient semiempirical approaches to the computation of electrostatic potentials, electric fields, and partial atomic charges have been presented. Although specifically designed for the implementation in combined quantum mechanical and molecular mechanical potentials, the more general formulation and the calibration of the model parameters against carefully chosen reference data allow application of the models in other areas of interest as well, among others in the definition of partial charges for molecular mechanics or molecular dynamics, and in the characterization of the electrostatic potential surfaces of large molecules.

Since the basic approximations of current semiempirical methodology also apply for our models, they are both easy to implement and numerically efficient. The expressions for the electrostatic potential and the electric field are based on AM1 and MNDO wave functions. They retain the

**FIGURE 8.**  $\alpha$ -D-idose and  $\beta$ -D-idose.

**TABLE XII.**  
**Atomic Partial Charges in  $\alpha$ -D-Idose and  $\beta$ -D-Idose.<sup>a-c</sup>**

Atom	6-31G* / PD <sup>d</sup>			QEq / PD		
	$\alpha$ -D-Idose	$\beta$ -D-Idose	Difference	$\alpha$ -D-Idose	$\beta$ -D-Idose	Difference
C-1	0.258	0.551	0.293	0.510	0.505	-0.005
C-2	0.199	-0.068	-0.267	0.163	0.123	-0.040
C-3	0.157	0.210	0.053	0.206	0.178	-0.028
C-4	0.036	0.062	0.026	0.170	0.172	0.002
C-5	0.286	0.286	0.000	0.205	0.186	-0.019
C-6	0.110	0.256	0.146	0.103	0.117	0.014
H-7	0.083	0.071	-0.012	0.009	0.042	0.033
H-8	0.061	0.105	0.045	0.009	0.011	0.002
H-9	0.067	0.079	0.012	0.017	0.034	0.017
H-10	0.075	0.068	-0.007	-0.005	0.001	0.006
H-11	0.127	0.078	-0.049	0.145	0.104	-0.041
H-12	0.019	-0.027	-0.046	0.025	0.026	0.001
H-13	0.063	0.028	-0.035	0.096	0.094	-0.002
O-14	-0.656	-0.741	-0.085	-0.657	-0.666	-0.009
O-15	-0.640	-0.620	0.020	-0.587	-0.593	-0.006
O-16	-0.699	-0.766	-0.067	-0.628	-0.669	-0.041
O-17	-0.609	-0.611	-0.002	-0.580	-0.582	-0.002
O-18	-0.400	-0.463	-0.063	-0.462	-0.450	0.012
O-19	-0.696	-0.721	-0.025	-0.610	-0.605	0.005
H-20	0.448	0.478	0.030	0.397	0.488	0.091
H-21	0.408	0.425	0.017	0.359	0.358	-0.001
H-22	0.449	0.469	0.020	0.372	0.374	0.002
H-23	0.412	0.408	-0.004	0.359	0.361	0.002
H-24	0.443	0.443	0.000	0.385	0.392	0.007

<sup>a</sup>All values in atomic units.<sup>b</sup>The values are based on geometries supplied by K. M. Merz, Jr.<sup>73</sup> which are identical to those used in ref. 72. The enumeration applied here and depicted in Figure 7 follows the conventions in the cited work.<sup>c</sup>The rms deviation with respect to 6-31G\* / PD:  $\alpha$ -D-Idose (0.078 e),  $\beta$ -D-Idose (0.080 e).<sup>d</sup>From ref. 72.

NDDO approximation and assume an orthogonal basis such that their computation affords negligible amounts of time. The evaluation of partial charges follows a simple scheme derived from Sanderson's electronegativity equalization principle and includes the Klopman-Ohno approximation for two-center interactions. It is basically a semiempirical and more economical reformulation of the QEq model proposed by Rappé and Goddard.

Reference data sets were consistently derived from *ab initio* RHF/6-31G\* data. For parameterization of the electrostatic potential and the electric field, *ab initio* potentials and fields on selected van der Waals surfaces were applied, while *ab initio* PD and Mulliken charges served a reference functions for the calibration of the charge equilibration model.

The final parameters allow us to compute electrostatic potentials and electric fields with a rel-

ative accuracy of about 20% and 60% (AM1), respectively, while the errors for MNDO wave functions are slightly larger (25% and 60%). The approach is similar to a model proposed by Ford and Wang, but it is calibrated against a much less restricted data set. As a consequence, it is considerably more accurate in regions near the molecular periphery and should thus be more appropriate for applications in combined QM/MM models.

The QEq model reproduces Mulliken and PD charges with an accuracy of about 0.05 e and 0.1 e, respectively. Although its performance is considerably better for Mulliken charges, much of the difference can be traced back to statistically ill-defined PD charges of buried atoms that constitute the molecular skeleton. With this in mind, and considering the success in reproducing most of the significant trends in homologous molecules, the results obtained with the QEq/PD model may also be regarded as satisfactory.

## Addendum

Another semiempirical charge model has recently been proposed.<sup>75</sup>

## Acknowledgments

This work has been supported in part through the Alfred Krupp Förderpreis. We wish to thank Professor K. M. Merz Jr. for providing us with the geometries of  $\alpha$ -D- and  $\beta$ -D-idose.

## Appendix

In this Appendix the analytical gradients of the geometry-dependent QEq charges are derived to yield computationally efficient formulas. These can easily be implemented in any code that performs geometry optimizations using energy formulas which rely on the proposed charge scheme.

We start with eq. (25) and differentiate the charge vector with respect to an arbitrary Cartesian coordinate  $x_I$  of atom  $I$ . Since the components of  $\mathbf{D}$  are constants, one immediately obtains

$$\mathbf{q}^{[I]} = \mathbf{C}^{-1[I]} \mathbf{D} \quad (\text{A.1})$$

Here we define the vector components  $q_K^{[I]}$  and the matrix elements  $C_{K,L}^{-1[I]}$  as follows:

$$q_K^{[I]} = \frac{\partial q_K}{\partial x_I} \quad (\text{A.2})$$

$$C_{K,L}^{-1[I]} = \frac{\partial C_{K,L}^{-1}}{\partial x_I} \quad (\text{A.3})$$

The matrix  $\mathbf{C}^{-1[I]}$  can easily be reduced to known or calculable quantities. Since the product of  $\mathbf{C}$  and its inverse yields the unit matrix, one readily verifies by differentiation:

$$\mathbf{C}^{-1[I]} = -\mathbf{C}^{-1} \mathbf{C}^{[I]} \mathbf{C}^{-1} \quad (\text{A.4})$$

$$\mathbf{q}^{[I]} = -\mathbf{C}^{-1} \mathbf{C}^{[I]} \mathbf{C}^{-1} \mathbf{D} \quad (\text{A.5})$$

We define the vector  $\mathbf{S}$  which is the same for any  $\mathbf{q}^{[I]}$ :

$$\mathbf{S} = \mathbf{C}^{-1} \mathbf{D} \quad (\text{A.6})$$

$$\mathbf{q}^{[I]} = -\mathbf{C}^{-1} \mathbf{C}^{[I]} \mathbf{S} \quad (\text{A.7})$$

According to eq. (A.7), two consecutive matrix operations of the type matrix times vector have to

be carried out for any Cartesian coordinate. This may be avoided if one recognizes that most of the derivatives of  $\mathbf{C}$  are zero. We define a new vector  $\mathbf{W}^{[I]}$

$$\mathbf{W}^{[I]} = \mathbf{C}^{[I]} \mathbf{S} \quad (\text{A.8})$$

the components of which may be derived from eq. (24):

$$W_K^{[I]} = J_{KI}^{[I]} S_I - J_{II}^{[I]} S_I \quad (K \neq I, I \neq 1)$$

$$W_I^{[I]} = -J_{II}^{[I]} S_I + \sum_{L=1}^n J_{IL}^{[I]} S_L \quad (I \neq 1)$$

$$W_K^{[1]} = J_{KI}^{[1]} S_1 - \sum_{L=2}^n J_{KL}^{[1]} S_L$$

$$W_1^{[1]} = 0 \quad (\text{A.9})$$

Using eqs. (26) and (27), the gradients of the two-center integrals

$$J_{KI}^{[I]} = \frac{\partial J_{KI}}{\partial x_I} \quad (\text{A.10})$$

are calculated to be

$$J_{IK}^{[I]} = e^2 (x_K - x_I) \left( r_{IK}^2 + \left( \frac{e^2}{2J_{II}} + \frac{e^2}{2J_{KK}} \right)^2 \right)^{-\frac{3}{2}} \quad (\text{A.11})$$

$$J_{IK}^{[I]} = J_{KI}^{[I]} = -J_{KI}^{[K]} = -J_{IK}^{[K]} \quad (\text{A.12})$$

$K$  is different from  $I$  for most of the vector components  $W_K^{[I]}$ . Thus only a few sums over all  $n$  atoms of the molecule need to be computed explicitly. The calculation of the derivatives of the QEq charges thus reduces to the simple multiplication of easily obtainable vectors  $\mathbf{W}^{[I]}$  with a matrix  $-\mathbf{C}^{-1}$  which is the same for any coordinate  $x_I$ :

$$\mathbf{q}^{[I]} = -\mathbf{C}^{-1} \mathbf{W}^{[I]} \quad (\text{A.13})$$

## References

1. A. Warshel and M. Levitt, *J. Mol. Biol.*, **103**, 227 (1976).
2. U. C. Singh and P. A. Kollman, *J. Comp. Chem.*, **7**, 718 (1986).
3. M. J. Field, P. A. Bash, and M. Karplus, *J. Comp. Chem.*, **11**, 700 (1990).
4. V. V. Vasilyev, A. A. Bliznyuk, and A. A. Voityuk, *Int. J. Quantum Chem.*, **44**, 897 (1992).
5. J. Åqvist and A. Warshel, *Chem. Rev.*, **93**, 2523 (1993).
6. V. Théry, D. Rinaldi, J.-L. Rivail, B. Maigret, and G. Ferenczy, *J. Comp. Chem.*, **15**, 269 (1994).
7. R. V. Stanton, D. S. Hartsough, and K. M. Merz, Jr., *J. Comp. Chem.*, **16**, 113 (1995).

8. P. A. Bash, M. J. Field, R. C. Davenport, G. A. Petsko, D. Ringe, and M. Karplus, *Biochemistry*, **30**, 5826 (1991).
9. V. V. Vasilyev, *J. Mol. Struct. (THEOCHEM)*, **304**, 129 (1994).
10. P. A. Bash, M. J. Field, and M. Karplus, *J. Am. Chem. Soc.*, **109**, 8092 (1987).
11. J. Gao, *J. Phys. Chem.*, **96**, 537, 6432 (1992).
12. J. Gao and J. J. Pavelites, *J. Am. Chem. Soc.*, **114**, 1912 (1992).
13. J. Gao and X. Xia, *Science*, **258**, 631 (1992).
14. J. Gao, *Int. J. Quantum Chem.: Quantum Chem. Symp.*, **27**, 491 (1993).
15. J. Gao, F. J. Luque, and M. Orozco, *J. Chem. Phys.*, **98**, 2975 (1993).
16. J. Gao, *J. Am. Chem. Soc.*, **115**, 2930 (1993).
17. J. Gao, *J. Am. Chem. Soc.*, **116**, 1563 (1994).
18. H. Liu, F. Müller-Plathe, and W. F. van Gunsteren, *J. Chem. Phys.*, **102**, 1722 (1995).
19. H. Liu and Y. Shi, *J. Comp. Chem.*, **15**, 1311 (1994).
20. D. Bakowies, Ph.D. thesis, Universität Zürich, Hartung-Gorre Verlag, Konstanz, 1994.
21. D. Bakowies and W. Thiel, to be published.
22. M. J. S. Dewar and W. Thiel, *J. Am. Chem. Soc.*, **99**, 4899 (1977).
23. M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, and J. J. P. Stewart, *J. Am. Chem. Soc.*, **107**, 3902 (1985).
24. N. L. Allinger, Y. H. Yuh, and R. Lii, program MM3(89), Technical Utilization Corporation, Powell, OH 43065, 1989.
25. N. L. Allinger, Y. H. Yuh, and J.-H. Lii, *J. Am. Chem. Soc.*, **111**, 8551 (1989).
26. J.-H. Lii and N. L. Allinger, *J. Am. Chem. Soc.*, **111**, 8566, 8576 (1989).
27. M. Scholz and H. J. Köhler, *Quantenchemie*, Vol. 3, Dr. Alfred Hüthig Verlag, Heidelberg, 1981.
28. C. Giessner-Prettre and A. Pullman, *Theor. Chim. Acta*, **25**, 83 (1972).
29. C. Alhambra, F. J. Luque, and M. Orozco, *J. Comp. Chem.*, **15**, 12 (1994).
30. P. L. Cummins and J. E. Gready, *Chem. Phys. Lett.*, **174**, 355 (1990).
31. C. A. Reynolds, G. G. Ferenczy, and W. G. Richards, *J. Mol. Struct. (THEOCHEM)*, **256**, 249 (1992).
32. G. G. Ferenczy, C. A. Reynolds, and W. G. Richards, *J. Comp. Chem.*, **11**, 159 (1990).
33. G. Rauhut and T. Clark, *J. Comp. Chem.*, **14**, 503 (1993).
34. B. Beck, G. Rauhut, and T. Clark, *J. Comp. Chem.*, **15**, 1064 (1994).
35. G. P. Ford and B. Wang, *J. Comp. Chem.*, **14**, 1101 (1993).
36. D. E. Williams, *Rev. Comp. Chem.*, **2**, 219 (1991).
37. R. T. Sanderson, *J. Am. Chem. Soc.*, **74**, 272 (1952).
38. R. T. Sanderson, *Science*, **121**, 207 (1955).
39. R. T. Sanderson, *Chemical Bonds and Bond Energy*, Academic Press, New York, 1976.
40. R. G. Parr, R. A. Donnelly, M. Levy, and W. E. Palke, *J. Chem. Phys.*, **68**, 3801 (1978).
41. J. Gasteiger and M. Marsili, *Tetrahedron*, **36**, 3219 (1980).
42. W. J. Mortier, S. K. Gosh, and S. Shankar, *J. Am. Chem. Soc.*, **108**, 4315 (1986).
43. A. K. Rappé and W. A. Goddard III, *J. Phys. Chem.*, **95**, 3358 (1991).
44. W. Thiel, program PP91, Universität Wuppertal, 1991.
45. P. K. Weiner, Ph.D. thesis, University of Texas at Austin, 1973.
46. M. Kolb, Ph.D. thesis, Universität Wuppertal, 1991.
47. W. J. Hehre, L. Radom, P. v. R. Schleyer, and J. A. Pople, *Ab Initio Molecular Orbital Theory*, John Wiley & Sons, Inc., New York, 1986.
48. M. L. Connolly, *J. Appl. Cryst.*, **16**, 548 (1983).
49. K. M. Merz, program SURFAC, University of California at San Francisco, 1989.
50. J. J. P. Stewart, program MOPAC 6.0, University of Indiana at Bloomington, QCPE No. 455, 1990.
51. B. H. Besler, K. M. Merz, and P. A. Kollman, *J. Comp. Chem.*, **11**, 431 (1990).
52. M. J. Frisch, G. W. Trucks, M. Head-Gordon, P. M. W. Gill, M. W. Wong, J. B. Foresman, B. G. Johnson, H. B. Schlegel, M. A. Robb, E. S. Replogle, R. Gomperts, J. L. Andres, K. Raghavachari, J. S. Binkley, C. Gonzalez, R. L. Martin, D. J. Fox, D. J. DeFrees, J. Baker, J. J. P. Stewart, and J. A. Pople, Gaussian 92, revision B, Gaussian Inc., Pittsburgh PA, 1992.
53. A. D. Buckingham, In *Intermolecular Interactions: From Diatomics to Biopolymers*, B. Pullmann, Ed., John Wiley & Sons, Chichester, 1978, p. 1.
54. F. J. Luque, F. Illas, and M. Orozco, *J. Comp. Chem.*, **11**, 416 (1990).
55. F. J. Luque and M. Orozco, *Chem. Phys. Lett.*, **168**, 269 (1990).
56. C. Alemán, F. J. Luque, and M. Orozco, *J. Comp. Chem.*, **14**, 799 (1993).
57. G. Klopman, *J. Am. Chem. Soc.*, **86**, 4550 (1964).
58. K. Ohno, *Theor. Chim. Acta*, **2**, 219 (1964).
59. B. Wang and G. P. Ford, *J. Comp. Chem.*, **15**, 200 (1994).
60. P. Politzer and H. Weinstein, *J. Chem. Phys.*, **71**, 4218 (1979).
61. R. S. Mulliken, *J. Chem. Phys.*, **2**, 782 (1934).
62. R. G. Parr and R. G. Pearson, *J. Am. Chem. Soc.*, **105**, 7512 (1983).
63. R. S. Mulliken, *J. Chem. Phys.*, **23**, 1833 (1955).
64. A. F. Reed, R. B. Weinstock, and F. Weinhold, *J. Chem. Phys.*, **83**, 735 (1985).
65. S. L. Price, R. J. Harrison, and M. F. Guest, *J. Comp. Chem.*, **10**, 552 (1989).
66. D. E. Williams, *J. Comp. Chem.*, **15**, 719 (1994).
67. L. Pauling, *The Nature of the Chemical Bond*, Oxford University Press, London, 1952.
68. C. I. Bayly, P. Cieplak, W. D. Cornell, and P. A. Kollman, *J. Phys. Chem.*, **97**, 10269 (1993).
69. W. D. Cornell, P. Cieplak, C. I. Bayly, and P. A. Kollman, *J. Am. Chem. Soc.*, **115**, 9620 (1993).
70. T. R. Stouch and D. E. Williams, *J. Comp. Chem.*, **14**, 858 (1993).
71. M. Orozco and F. J. Luque, *J. Comp. Chem.*, **11**, 909 (1990).
72. K. M. Merz, Jr., *J. Comp. Chem.*, **13**, 749 (1992).
73. K. M. Merz, Jr., private communication, 1994.
74. *CRC Handbook of Chemistry and Physics*, CRC Press, Boca Raton, FL, 1992.
75. J. W. Storer, D. J. Giesen, C. J. Cramer, and D. G. Truhlar, *J. Comput.-Aided Mol. Design*, **9**, 87 (1995).